



12-2008

The Consequential Effects of High-Stakes Testing on Teacher Pedagogy, Practice and Identity: Teacher Voices Disrupt the A Priori

Janice Ellen Blake
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss



Part of the [Education Commons](#)

Recommended Citation

Blake, Janice Ellen, "The Consequential Effects of High-Stakes Testing on Teacher Pedagogy, Practice and Identity: Teacher Voices Disrupt the A Priori. " PhD diss., University of Tennessee, 2008.
https://trace.tennessee.edu/utk_graddiss/476

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Janice Ellen Blake entitled "The Consequential Effects of High-Stakes Testing on Teacher Pedagogy, Practice and Identity: Teacher Voices Disrupt the A Priori." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Richard L. Allington, Major Professor

We have read this dissertation and recommend its acceptance:

Anne McGill-Franzen, Gary Skolits, Stergios Botzakis

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Janice Ellen Blake entitled “The Consequential Effects of High-Stakes Testing on Teacher Pedagogy, Practice and Identity: Teacher Voices Disrupt the A Priori.” I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Richard L. Allington, Major Professor

We have read this dissertation
and recommend its acceptance:

Anne McGill-Franzen

Gary Skolits

Stergios Botzakis

Accepted for the Council:

Carolyn R. Hodges, Vice Provost and
Dean of the Graduate School

(Original signatures are on file with official student records.)

THE CONSEQUENTIAL EFFECTS OF HIGH-STAKES TESTING ON
TEACHER PEDAGOGY, PRACTICE AND IDENTITY: TEACHER
VOICES DISRUPT THE A PRIORI

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Janice Ellen Blake
December 2008

Copyright © 2008 by Janice E. Blake
All rights reserved.

DEDICATION

This work is dedicated to the important people in my life:

To my mother, Irene McDonald, who has guided me through all the days of my life.

To my dad, William McDonald, who reminded me of the importance of “the direction we are moving”.

To our children, Aiden and Reghan, who are nothing less than “everything”.

To my husband, Verdie, who continues to be my greatest supporter and the love of my life.

To my family and my friends, how lucky are we to be a part of great things, great love, great moments of glorious wonder and those essential moments of great ridiculousness.

Thank you.

ACKNOWLEDGEMENTS

Several people deserve special thanks for their contribution and support of my research:

First, I thank Dr. Richard Allington, who, before I came to Tennessee, was just “words on a page” that guided some of my most challenging days as a teacher in a small school district in Canada. His pragmatic writing, sensible thinking, and just plain good sense brought this teacher on a long journey to Tennessee from Beautiful British Columbia, Canada.

To my dissertation committee: I have wondered if each of you were aware of your profound effect on my work, my confidence, and my success. Dr. Anne McGill-Franzen was the true unfound “gem” for me – you are brilliant, thoughtful and always right on. Dr. Skolits I want to give you a rock – as this is how I will always think of you; kind and solid with just the right amount of very appreciated humor. Dr. Botzakis, I hope I have carried you with me on this survey journey – as it turned out it really was so much more than – just a survey.

To Patti Fagg and Vicki Church: You are both such very special, supportive people. My thanks for listening, laughing and always telling me, “You can do it”.

And to all of my colleagues: It is so true that we need each other as friends, guides, tutors and so much more along this challenging journey. Who would have ever known that on our way to become researchers we would also become some of the greatest cheerleaders ever! We all have agreed that know one would understand just how challenging this journey has been but each of you.

To the teachers,

"There's no word in the language I revere more than 'teacher.' My heart sings when a kid refers to me as his teacher, and it always has. I've honored myself and the entire family of man by becoming a teacher."
(Pat Conroy, writing as the main character in *Prince of Tides*)

ABSTRACT

Today, under the federal mandates of No Child Left Behind (NCLB), test scores are being used for ways and means in which they were never designed, normed or intended (Linn, 2003). As a result, the purposes and uses of high-stakes tests have become a source of concerned debate among stakeholders, who see the consequences of high-stakes testing as having significant effects within the larger educational reform known as No Child Left Behind (NCLB) (Amrein & Berliner, 2002b). Allington (2002) has stated that NCLB has dramatically changed the testing story, making high-stakes tests one of the leading and central characters of the current reform. Previous research of high-stakes testing has tended to exclude the voice of those closest to the issues and concerns – the teacher. Utilizing quantitative survey methodology, two central research questions guided this research, asking:

1. What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?
2. What are the consequential effects of high-stakes testing in relation to teachers' work and identity?

This study examined the perceptions of teachers currently working within the high-stakes testing environment in Southeastern Tennessee. A review of the literature is presented, as well as results from a 63-item survey of teachers. Analyses of these data reveal that high-stakes testing does indeed affect teacher pedagogy, practice and identity in highly unfavorable ways. Results from this study represent 408 teachers responding to

the survey instrument. Additionally, 125 teachers responded to an optional open-ended text question reporting that high-stakes tests both influence and impact instruction and most importantly contradicts teachers' views of sound educational practice. Results indicated that elementary teachers teaching in below average performing schools situated in rural areas are the most profoundly impacted by high-stakes testing.

TABLE OF CONTENTS

CHAPTER 1	1
Chapter Introduction	1
High-Stakes Testing in the Twenty-First Century	2
Statement of the Problem	5
Historical Context of High-Stakes Testing	12
Defining High-Stakes Testing	14
Defining High-Stakes Testing: The Construct	14
Defining High-Stakes Testing: Educational Research	16
Defining High-Stakes Testing: Professional Research Organizations	17
Situated Context of Study	20
Theoretical Framework and Conceptual Apparatus	25
Purpose of the Study	30
Research Questions	31
Significance of the Study	32
Assumptions and Researcher Ideological Stance	35
Limitations of the Study	39
Delimitations of the Study	40
Definition of Terms	40
Organization of the Study	44
CHAPTER 2 REVIEW OF THE LITERATURE	46
Chapter Introduction	46
Teacher Identity Framed	48
Identity and Process	51
Identity and Commitment	54
Identity and the Given and the Possible	57
Identity and Becoming a Teacher	63
Identity and Courage	67
Identity and a High-Stakes Test	70
Identity in Review	73
High-Stakes Testing through Diverse Lenses	76
High-Stakes Testing: Proponent and Opponent Views	77
High-Stakes Testing: Methodological Diversity	82
Curricular Control: Metasynthesis	83
Instructional Influence: Case Study	84
Student Proficiency and Outcomes: Time Analysis	87
High-Stakes and Reform: Mixed Methods	89
Proponent, Opponent and Diversity in Review	92
High-Stakes Testing and Quantitative Survey Methodology	93
Survey Research and High-Stakes Testing	94
Quantitative Survey: Corruption, Impact and Perceptions	96
Quantitative Survey: “Voices” from the Frontlines	100

Quantitative Survey: Mediating High-Stakes Reform.....	102
Quantitative Survey: Impact and Instructional Practices.....	104
Quantitative Survey: Negotiating What’s In and Out.....	107
Quantitative Survey: Surveying the Nation’s Teachers.....	109
Quantitative Survey Methodology in Review.....	113
Literature Review in Summary	114
CHAPTER 3 METHODOLOGY	117
Chapter Introduction	117
Rationale	117
Survey as “Best Fit”	117
Survey as “Logistics in Use”	121
Research Questions:	122
Research Design.....	123
Instrumentation	125
Modeling Upon a Nationally Normed Survey	127
Construct Validity	128
Survey Instrument.....	129
Survey Domains.....	131
Sampling Frame	133
Sample Size.....	135
Data Collection	137
Data Analysis.....	142
Data Analysis: Survey Instrument	143
Data Analysis: Narrative Data	144
Methodology Conclusion.....	150
Methodology Summary	154
CHAPTER 4 RESULTS	155
Chapter Introduction	155
Overview of the Sample	157
Demographic Frequencies	157
High-Stakes Tests and Practice	161
High-Stakes Testing and Hours of Preparation	162
High-Stakes Testing and Frequency of Preparation	162
High-Stakes Tests and Alignment.....	164
Reliability and Validity of Data	164
Factor 1: Value.....	167
Factor 2: Pressure.....	169
Factor 3: Climate.....	169
Factor 4: Alignment	171
Results of Primary Research Question	173
High-Stakes Testing: Impact on Perceptions.....	174
Perception Difference by Independent Variables	174
Perception Difference by Grade	175

Perception Difference by Setting.....	176
Perception Difference by Experience	177
Perception Difference by Performance	178
High-Stakes Testing: Impact on Pedagogy and Practice	181
Impact of Preparation by Grade	182
Impact of Preparation by Setting	184
Impact of Preparation by Experience.....	186
Impact of Preparation by Performance	186
Descriptive Statistics—Instructional Time	187
Instructional Time by Grade	188
Instructional Time by Setting.....	190
Instructional Time by Experience	190
Instructional Time by Performance	190
Descriptive Statistics—Modes of Instruction	196
Modes of Instruction by Grade	197
Modes of Instruction by Setting.....	197
Modes of Instruction by Experience	197
Modes of Instruction by Performance.....	198
Open-Ended Survey Question: Narrative Data.....	198
Test Preparation and Curriculum Narrowing.....	199
Teacher Identity.....	202
Test Use and Value of Test	203
Sociocultural Influences and Intervening Variables.....	204
Rewards and Sanctions.....	206
Proponent and opponent view.....	208
Connectedness of the effects of high-stakes testing	209
Results of Analysis	211
Results Summary	211
CHAPTER 5 DISCUSSION.....	213
Chapter Introduction	213
Review of Research Rigor	214
Discussion.....	215
Research Questions.....	217
Recommendations.....	245
LIST OF REFERENCES	253
APPENDICES	276
APPENDIX A: TEACHER SURVEY ON THE IMPACT OF STATE-MANDATED TESTING PROGRAMS	277
APPENDIX B: TEACHER SURVEY ON THE EFFECTS OF HIGH-STAKES TESTING ON TEACHER PEDAGOGY, PRACTICE AND IDENTITY	286
APPENDIX C: INITIAL FACTOR LOADINGS	301
APPENDIX D: FINAL FACTOR LOADINGS.....	304
VITA	307

LIST OF TABLES

Table 3:01: Pedulla et al. study reliability	128
Table 4:01: Frequencies for school setting	158
Table 4:02: Frequencies for grade	159
Table 4:03: Frequencies of grade level grouping	159
Table 4:04: Frequencies for teaching experience	160
Table 4:05: Frequencies for school designation	161
Table 4:06: Frequencies for Title 1 and AYP	161
Table 4:07: State-mandated test influence on teaching	162
Table 4:08: Class hours per year spent preparing students for the state-mandated test	163
Table 4:09: Test preparation specifically for the state-mandated test during school year	163
Table 4:10: Alignment of test preparation materials to content of state-mandated test .	165
Table 4.11: Factor loadings for value	168
Table 4:12: Factor loadings for pressure	170
Table 4:13: Factor loadings for climate	171
Table 4:14: Factor loadings for alignment.....	172
Table 4.15: Reliabilities	172
Table 4:16: Impact on perception	174
Table 4:17: Perception difference by grade	175
Table 4:18: Perception difference by setting	177
Table 4:19: Perception difference by experience.....	178
Table 4:20: Perception difference by performance.....	179
Table 4:21: Preparation: impact on pedagogy and practice.....	181
Table 4:22: Impact of preparation by grade.....	182
Table 4:23: Impact of preparation by setting.....	184
Table 4:24: Impact of preparation by performance	187
Table 4:25: Instructional time	189
Table 4:26: Instructional time by grade	191
Table 4:27: Instructional time by performance.....	192
Table 4:28: Modes of instruction.....	196

LIST OF FIGURES

Figure 1:01: Theoretical framework and conceptual apparatus.....	26
Figure 4:01: Connectedness of the effects of high-stakes testing.....	210

CHAPTER 1

“We are entering the age of infinite examination and of compulsory objectification” (Foucault, 1995, p. 189).

Chapter Introduction

Testing tied to educational reform, as we know it today, had its inception in the mid-1930s. By the end of the 1940s almost every school in the United States was using some form of standardized test (Callahan, 1962). During the 1950s tracking and selection were at the forefront of educational goals under the implementation of the National Defense Education Act (NDEA). The 1960s saw a move to utilize test scores for program accountability of high-poverty schools with the implementation of the Elementary and Secondary Education Act (ESEA) and began a massive increase in testing. The 1970s marked the beginning of state mandated minimum competency testing. A sharp move in the 1980s saw randomly sampled test scores being flagged for purposes of identifying overall school system accountability under the initiatives of the National Assessment of Educational Progress (NAEP) report. The federal standards-based accountability movement became the hallmark of the 1990s as proficiency standards were mandated and then acted as the benchmark of testing and primary indicator of student knowledge, teacher proficiency and school effectiveness (Cross, 2004).

Today, under the federal mandates of No Child Left Behind (NCLB), test scores are being used for ways and means in which they were never designed, normed or

intended (Linn, 2003). In the half century since Sputnik, teachers have weathered a plethora of federal, state and local educational reform. “Over the last 15 years, the movement for higher standards and accountability in our schools has led several states – and now the federal government, with the No Child Left Behind (NCLB) Act – to adopt test-based accountability policies” (Goldberg, 2004). The last half century has seen consequential shifts in educational policies and the implementation of far-reaching reform mandates resulting from the current NCLB Act.

High-Stakes Testing in the Twenty-First Century

High-stakes testing has become a central national policy issue (Allington, 2002; Cuban, 2007) where national and state policies within the reform agenda are propelled and supported by mandated high-stakes tests. These high-stakes tests are seemingly held as a metaphor for standards of student learning, teacher effectiveness, and school success, where a number or score shapes our perceptions and our objectivity within the educational environment (Dorn, 1998). With such sweeping expectation, meaning and consequence attached to a single test score one is left to question a reform which, seemingly, is antithetical to the purpose of education (Dewey, 1938).

Nichols and Berliner (2007) ask us to consider how and why high-stakes testing has so seamlessly slipped into the culture of education in America. Further to this perspective, Stone (2002) cautions the overreliance on a single number reminding policymakers that “to select one feature of something, assert a likeness on the basis of that feature, and ignore all other features” (p. 165) is to exclude or ignore what may the

most essential to educational reform. Here, high-stakes tests present as a complex agenda driven force behind the current standardized testing mandates.

As a result, the purposes and uses of high-stakes tests have become a source of concerned debate among stakeholders, who see the consequences of high-stakes testing as having significant effects within the larger educational reform known as No Child Left Behind (NCLB) (Amrein & Berliner, 2002b). Aspects of the NCLB Act form a core of accountability where the intent of the Act places an emphasis on standards of learning for all children and directs greater attention to those groups of children who have been largely ignored or marginalized in the past. While the NCLB Act “stays the course of standards-based reform and encourages states to adopt ambitious subject-matter standards” (Linn, 2003, p. 4), exactly how states have mandated and enacted these reforms have played out and resulted in highly test responsive and test dominated educational settings. As a result, the compliance and reliance on high-testing has co-opted the intent and promise of the NCLB Act.

Allington (2002) stated that NCLB has dramatically changed the testing story, making high-stakes tests one of the leading and central characters of the current reform. Amidst the persistent concerns focusing on determining and reporting the quality of education and performance of schools and teachers, high-stakes testing is viewed as a decidedly efficient way to obtain numbers and scores, which are then directly transferred to highly publicized standards measures (Kohn, 2000; Kozol, 2005; Nichols & Berliner, 2007). In contrast to the perspective that test scores are both reliable and valid measures

of achievement, Heubert and Hauser (1999) contend that standardized group achievement tests are neither.

The over simplification of tests and accountability (Linn, 2003) appears to be driving much of the current educational purpose, pushing schools towards ill-conceived standards and accountability agendas (Spillane, 2004) where high-stakes testing is at the center. Further to this perspective, Stone (2002) describes the outcomes resulting from high-stakes testing as one of striving towards the lowest cost objectives where “[g]etting the most out of a given input or achieving an objective for the lowest cost are simple definitions of the goal of efficiency” (p. 61). The effects of high-stakes testing as an accountability tool is systematically influencing the standards reform, thereby resulting in a high-stakes testing environment which ultimately both influences and impacts teacher pedagogy, practice, and identity.

Historically, high-stakes tests have proven to be efficient tools in the production of numbers and scores, but conversely have presented as highly unstable instruments and indicators of effective teaching and student learning (Linn, 2000). Allington (2002) argues that teachers who are caught within the policy trap with “less and less professional autonomy paired with more and more accountability” (p. 33) find themselves losing more and more of their teacher professional autonomy as they work within the high-stakes test environment and the production of numbers and scores. While the initial intent of NCLB was to set educational standards, improve the educational learning opportunities for all students and thereby raise achievement scores of students, the current outcomes of NCLB

appear to have cast a normalizing gaze, homogenizing and mandating a standardized and highly test based educational setting focused on achievement and conformity.

This study draws upon teacher expertise and experience rooted within the current everyday terrain of high-stakes testing (Gardiner, 2004) to document teachers lived experience within the high-stakes testing environment (Pedulla et al., 2003). Standards, accountability and testing are central to the educational reform movement. What are the consequential effects of high-stakes testing on teachers' pedagogy, practice and identity? How do teachers position their teacher identity within the high-stakes testing environment? These are the questions asked by teachers and researchers (Barksdale-Ladd & Thomas, 2000; Linn, 2003) in the field, and which delineate and describe the central purpose of this study. I am interested in how teachers interpret and mediate the high-stakes testing mandates in the active pedagogical contexts and dimensions of their classroom practices and teacher identity. This study recognizes the importance of teacher voice within the current reform movement.

Statement of the Problem

This study investigates teacher reports of the effects of high-stakes testing on teacher pedagogy, practice, and identity. Caught in the current press to evaluate, measure and report on student achievement, teachers are finding it increasingly challenging to attend to the job of teaching the students they have in their classrooms (Gunzenhauser, 2003). High-stakes testing has become the *moving* bar to which teachers must direct their daily teaching and attention (Riddle, Buly & Valencia, 2002). State curriculum

frameworks, content standards, and mandated assessment programs to measure student achievement describe the momentum of the reform movement over the last decade (Hoffman, Assaf, & Paris, 2001; Pedulla et al., 2003). With the current mandates of NCLB calling for 100 percent proficiency of the nation's children in reading and mathematics by the year 2014, the stakes are high. High-stakes testing is based on the beliefs stemming from the NCLB Act that consequences attached to accountability measures are effective and reliable ways to motivate teachers and to ensure higher student academic performance (Allington, 2003).

Standards and accountability are the central components of the current American education reform movement. Over the last 50 years the United States education system has experienced waves of educational reform (Linn, 2000), where high-stakes tests have become increasingly viewed as a lever for educational change. Sorting by means of a test has become the norm, thereby providing a symbol of a standardized expectation supported by a metaphorical means to judge and compare achievement (Mehan, 1990). As a result, teachers are experiencing the far-reaching effects of the high-stakes educational environment, impacting their pedagogy, practice and teacher identity (Hilliard, 2000; Hoffman et al., 2001). "Reform by comparison" (Corbett & Wilson, 1991, p. 2) describes the metaphorical accountability frame (Lakoff & Johnson, 1980) where high-stakes testing has become a standard of "proficiency as illusion" (Cronin, Dahlin, Adkins, & Kingsbury, 2007) within the current standards movement.

Quite simply, it takes times to think about complex issues. In the current press to assess, evaluate and account for student achievement and teacher effectiveness we are in danger of trying to think too hard and too quickly when it comes to “failure” (Allington, 1994a). Teachers who work in high-stakes subject areas and grades must accomplish what all other previous reform efforts have failed to do - close the achievement gap for all children (Abrams, Pedulla, & Madaus, 2003).

Under federal mandates, states are individually setting the “bar” and periodically raising achievement thresholds, until the 100 percent target performance level for all children is met in 2014. Consequently, in this era of accountability and data-driven decision making (Individuals with Disabilities Education Act, No Child Left Behind Act), teachers are caught in the press and held accountable to raise student achievement and test scores utilizing highly unstable high-stakes accountability measures (McGill-Franzen & Allington, 2006; Thomas, 2005). The resulting mandated policies, measurement-driven curriculum and instruction have produced both intended and unintended consequences for educators.

The current educational climate of high-stakes testing demonstrates the tenuous nature of the reform movement from its inception with *A Nation at Risk* (1983) to the NCLB (2002) mandates of standardization and accountability. *A Nation at Risk* called for "an end to the minimum competency testing movement and the beginning of a high-stakes testing movement that would raise the nation's standards of achievement drastically" (Amrein & Berliner, 2002a, p. 4). The results of this historic and politically

driven report (Amrein & Berliner, 2002a; Berliner & Biddle, 1995) became the persuasive seeds of the NCLB Act, where the driving force in education today has become the high-stakes accountability measures mandated by federal law and overseen by governmental agents. As a result, as “the bar” is raised and standards are set at the state level (Riddle Buly & Valencia, 2002), teachers report that they are overwhelmed and express they are “failing their kids” (Johnson, 2002, p. 1).

Viewing the historical path of high-stakes testing from this perspective, it appears teachers have been led to a kind of unvoiced inductive fallacy of complicity, compliance and complacency (Nichols & Berliner, 2007). The federally instigated press to ensure high standards and increased student achievement has become the central assertion for the utilization of high-stakes accountability measures. The promise of these reform measures has played out as a kind of ‘silver bullet’ offering to ensure that schools move away from the highly publicized downward spiral of American educational standards and achievement (Guthrie & Springer, 2004).

Over the last decade, as states have increasingly come to rely on high-stakes assessments to determine the success and effectiveness of educational settings, teachers have been held to high-stakes accountability measures in which they have little opportunity for input or control. Heightened public attention to a single test score, along with federal mandates to move education from the perceived low standards of achievement and effectiveness, has created an arduous situation for the work of teachers.

Reporting results from a large scale reform initiative in Kentucky, Datnow, Hubbard and Mehan (2002) stated that the “high stakes accountability system – a powerful external structure – had the unintended consequences of destabilizing reform efforts in most of the schools” (p. 133) where teachers were often confused, and sidelined any and all local reform initiatives in the efforts to prepare and align their teaching practices to the test. Recent findings show, at the onset of large scale high-stakes reforms, teachers working within the high-stakes testing environment find themselves largely the focus of the success or failure of the current educational accountability (Linn, 2003) with little to no input. Such a counter intuitive pedagogical and practice choice runs antithetically against the stated intent and purpose of the high-stakes testing mandates in schools across the nation.

More recently, in an introduction to the Cronin et al (2007) report, Finn and Petrilli (2007) acknowledged the importance of sound standards as the primary foundation for all aspects of the reform agenda, stating that:

Standards-based reform hinges on the assumption that one can trust the standards, that they are stable anchors to which the educational accountability vessel is moored. If the anchor doesn’t hold firm, the vessel moves – and if the anchor really slips, the vessel can crash against the rocks or be lost at sea (p. 4).

This current statement highlights the importance of clarity, practical knowledge, and trust within any standards and accountability reform movement and highlights the importance of teacher voice and participation within the conversation. In this respect, Cuban (2007) contended that the “paradox of distrusting teachers and then turning around and expecting

them to solve the problems of low-performing students has often frustrated critics and reformers” (p. 2) and teachers.

Currently, high-stakes testing is utilized for such far-reaching accountability outcomes that the effects of this phenomenon must be understood from the perspective voices of the teachers who are influenced and impacted within broad, highly variable and non-standardized educational settings. While researchers have affirmed the importance of teacher voice within the overall effectiveness of any reform movement, educational policy and mandates continue to place the teacher as purveyor of the high-stakes testing mandates and not “regarded as knowledgeable agents in the debate” (Smith & Fey, 2000, p. 343). Finn and Petrilli’s metaphor calls for a reconceptualization of who is included in these important conversations and invites the voice of teachers to be valued and included articulately so as to ensure that policy-making represents teachers’ practical knowledge and experiences.

In one of the broadest national surveys conducted on these issues in the United States, Pedulla, Abrams, Madaus, Russell, Ramos, and Mia (2003) found that the current state testing initiatives compelled teachers to change their pedagogy and practice so as to comply with standards and accountability mandates. The conclusions of the Pedulla et al. (2003) study calls for teachers to be included as an essential and integral voice within the reform agenda. “Teachers are on the front line every day,” said Joseph Pedulla. ... Their voice on this issue must be heard; their opinions must enter into the formation of sound

testing policy” (p. 2). Paradoxically, within the formation of educational policy, teacher voices are rarely heard in the testing debate.

At the heart of this problem is the current emphasis on high-stakes testing which has caused a highly test responsive teaching and learning environment. Throughout the profession, teachers under the “gun” of the No Child Left Behind Act (NCLB, 2002) are finding that these high-stakes tests hold consequences for both teaching and learning (Johnston, 1998; Linn, 2003; Messick, 1989; Shepard, 2000). More recently, Valli, Croninger, Chamblis, Graeber and Buese (2008) stated “a central paradox of NCLB: as more emphasis is placed on assessment results, particularly in the form of higher standardized-test scores, less emphasis is placed on professional standards for teaching and learning” (p. 3). For example, Duffy (2007) described the tenuous high-stakes testing environment where teachers find themselves mandated to deliver highly scripted curriculum, utilizing prescribed teaching methods, emphasizing test materials and low-level test preparation types of daily instruction where the primary goal of *teaching* is “to get students to score well on end-of-year tests” (p. 7).

“Classroom stories and teacher surveys report again and again that more lesson time is spent preparing students for high-stakes tests and the narrowing of the curriculum to what is on those tests” (Cuban, 2007, p. 14). High-stakes testing is a highly debated contemporary issue which is politically charged and framed as a “scientific” approach to monitoring school productivity. Student learning and achievement reported in single raw

scores and percentile rankings, elevates the function and utility of high-stakes testing to a somewhat mythical realm.

Based on the beliefs stemming from the NCLB Act, the resulting consequences attached to high-stakes accountability measures are viewed by policymakers as effective and reliable ways to motivate teachers and to ensure higher student academic performance. This rationale, rooted in presumptions that high-stakes tests will raise the standard of teacher pedagogy and practice, promotes a set of conditions which encourages a culture of mistrust and inflated expectations (Allington, 2003) within the current reform agenda. In response, researchers (Abrams et al., 2003; Cohn & Kottkamp, 1993; Lortie, 1975; Pedulla et al., 2003) have argued for the inclusion of the collective teacher voice in future educational policy decisions and implementation. Therefore, critical to developing an understanding of the teacher experience in terms of high-stakes testing in these educational times, is capturing and documenting the experienced voice of the respondents.

Historical Context of High-Stakes Testing

Student achievement has become the prevailing concern in education. In 1983, *A Nation at Risk: The Imperative for Educational Reform* (NCEE) was published in the United States, declaring that students in American schools were failing. The result of this publication was the heightened scrutiny of curriculum, instruction, teachers, schools, and ultimately set in motion the current reform movement in which high-stakes testing today is at its center. Under the Reagan administration, the federal government began to

encourage states to "raise standards, increase testing, establish accountability, strengthen requirements, and secure better teachers and better teaching" (Perkinson, 1995, p. 370). School reform was given its charge and the monitor and gate-keeper became high-stakes testing as we know it today.

This single report (NCEE, 1983) moved the education agenda from teaching the basics to monitoring school quality and setting standards of performance and proficiency (McGill-Franzen, 2000). Accountability, in the form of high-stakes tests at all levels, was seen as the key to raising individual student proficiencies. The pressure cooker of the high-stakes testing education environment was just beginning to come to a boil. Report cards of student test performance were now the standards to determine how and where schools and school systems placed on a national comparison. Education became a footrace where educators were now under considerable pressure of high-stakes consequences to raise test scores. Overwhelmingly, teachers responded to this high-stakes teaching environment by focusing on the skills being tested, teaching test-taking skills, using test format to guide and form their daily teaching content and practices (Allington, 2002; Amrein & Berliner, 2002b; Corbett & Wilson, 1991; McGill-Franzen & Allington, 2006). High-stakes testing became a polished machine reporting student achievement. The serious issue within education research and practice has become establishing a common understanding of what *high-stakes testing* within the initiatives of reform means to all constituents - this is a critical element that continues to remain undefined.

Defining High-Stakes Testing

Defining High-Stakes Testing: The Construct

Although the term *high-stakes* testing is utilized and frequently referred to in the current research, few educational researchers specifically state a definition of high-stakes testing in the presentation of their research. Yet current literature is found to be replete with references highlighting and debating the effects of high-stakes testing. A review of the literature found that few research studies specifically define the construct. In addition, this review determined that while educational researchers have made less of a specific definition explicit within their writing, educational theorists, writers, and journalists have frequently made assertions referencing and defining high-stakes testing in the literature (Bracey, 2000; Kohn, 2000; Mehrens & Popham, 1992; Nichols & Berliner, 2007; Resnick & Resnick, 1985). These writers have defined high-stakes testing as being both problematic and spurious as well as being instruments of sound educational policy.

Those researchers who have stated definitions of high-stakes testing in their scholarly writing generally define the construct in markedly similar ways. According to Spring (2004), high-stakes testing “refers to an examination that determines a person’s future academic career and job opportunities” (p. 36). Crawford and Impara (2001) stated that “(W)henever assessments affect the lives of students, we may consider those to be high-stakes tests” (p. 140). Franzak (2004) described high-stake tests, stating “(S)tandardized assessments become high-stakes when educational or personnel decisions are based on the results” (p. 235). Drawing from their research, Paris, Lawton,

Turner and Roth (1991) defined high-stakes tests as those tests where “the consequences are profound for the respondents” (p. 12). Jones and her colleagues (Jones, Jones, & Hargrove, 2003) argued that tests acting as a “way to measure student achievement and school quality and as a mechanism to hold students and educators accountable” (p. 1) are held as high-stakes test. In a recent chapter in *What Research Has To say About Reading Instruction*, Guthrie (2002) stated that “(A) test or testing program is called high-stakes when it is used to make important decisions about individual students, teachers, or schools” (p. 370). Taken as a whole, this educational literature draws upon the historical and sociocultural context and describes the prevailing definitions of high-stakes testing, thereby representing a composite definition which is reflective of theory, research and practice.

Within the educational research forum Madaus’ (1988a) highly referenced (Au, 2007; Cimbricz, 2002; Grant, 2000; Heubert & Hauser, 1999; Horn, 2003; Mathison & Freeman, 2003; Rex & Nelson, 2004) definition of high-stakes testing offers educational researchers a sound description of the phenomenon. Madaus’ definition represents a construct which many researchers, studying mandated testing, have embraced as describing the definition of “high-stakes testing” in educational research:

High-stakes tests include those used for the certification or recertification of teachers, promotion of students from one grade to the next, award of a high school diploma, assignment of a student to a remedial class, allocation of funds to a school or school district, award of merit pay to teachers on the basis of their students’ test performance, certification or recertification of a school district, and placement of a school system in “educational receivership”. (p. 30)

Educational researchers who advocate that high-stakes testing holds great potential to both monitor and increase student achievement share a definition of high-stakes testing as one which has the ability and purpose to act as a lever of change within a contemporary educational reform movement (Grant, 2000). Consequently, the interpretations of a commonly held definition of high-stakes testing has become somewhat of a politically charged and divisive construct, where definitions are manipulated and blurred. With high-stake tests ultimately and profoundly influencing peoples' lives (Downing & Haladyna, 1996) outside of the stated NCLB intent of accountability and responsibility it is essential that a common definition is shared among all stakeholders.

Defining High-Stakes Testing: Educational Research

While many scholars and researchers currently debate the definition and significance of high-stakes testing, the utilization of large-scale high-stakes tests have become a distinct piece of the educational landscape over the last 30 years (Heubert & Hauser, 1999). Drawing upon research of statewide testing programs in Pennsylvania and Maryland, Heubert and Hauser concluded that the perceived definition and the level of stakes associated with tests were less characteristics of the test, per se, but rather greater characteristics of the perceptions of test use. These researchers asserted that the use of a single indicator of student learning to make high-stakes decisions about tracking, promotions and graduation was unethical. Heubert and Hauser (1999) contend that “people may attach a level of stakes to a test that is out of character with the formal consequences associated with it” (p. 26) suggesting that the importance of a commonly

held definition of *high-stakes* may be more readily understood and defined at the local level rather than by federal mandates.

However, educational researchers with opposing perspectives see this definition of high-stakes testing as problematic; in contrast, they define high-stakes testing as holding major consequences for students, teachers, and schools - calling for a clear understanding of the intent and outcome of President Bush's original initiative of NCLB. This politicization of assessment and accountability is described by Hillocks (2002) in his landmark study of how state assessments control learning. Hillocks defined high-stakes testing as assessments where "the fortunes of individual students, schools, and school districts rise or fall on the results" (p. 18). Johnson and Johnson (2006) published an in-depth study of poverty, testing and failure, asserting that high-stakes tests are those which "base life-altering decisions of single test scores" (p. 202).

For these educational researchers high-stakes tests are those tests which critically impact programs, curriculum, and individual student achievement resulting in high-stakes consequences within the educational setting. While researchers may hold specific characteristics of the definition of high-stakes testing in contrast to others, it is essential that the commonly held definition is reflective of current educational research and educational practice, supporting a common construct and purpose of "high-stakes" testing.

Defining High-Stakes Testing: Professional Research Organizations

Professional educational research organizations, recognizing the need to define and state their organizational position regarding high-stakes testing, have issued position

papers defining high-stakes testing. Drawing from the 1999 *Standards for Educational Psychological Testing*, the American Educational Research Association (AERA, 1999) stated, high-stakes test are those which:

carry serious consequences for students or for educators. Schools may be judged according to the school-wide average scores of their students. High school-wide scores may bring public praise or financial rewards; low scores may bring public embarrassment or heavy sanctions. For individual students, high scores may bring a special diploma attesting to exceptional academic accomplishment; low scores may result in students being held back in grade or denied a high school diploma (p.1).

The National Reading Conference (NRC, 2004) published a comprehensive policy brief “focusing on the popularity of high-stakes tests, the uses and misuses of high-stakes tests and the consequences of high-stakes testing” (p.2). In this policy brief, Afferbach states that:

[h]igh-stakes (reading) tests are those with highly consequential outcomes for students, teachers, and schools. These outcomes may include promotion or retention, student placement in (reading) groups, school funding decisions, labeling of schools as successful or failing and the degree of community support for a school. (p. 2)

Additionally, the International Reading Association (IRA, 1999) issued a position paper stating:

[h]igh-stakes testing means that the consequences for good (high) or poor (low) performance on a test are substantial. In other words, some very important decisions, such as promotion or retention, entrance into an educational institution, teacher salary, or a school district’s autonomy depend on a single test score. (p. 2)

Equally, these statements from major educational research organizations present a common, defining construct of high-stakes testing within the educational research community.

Defining High-Stakes Testing: Study Definition

While high-stakes testing continues to be a highly complex and multi-faceted construct, *high-stakes testing* in educational research refers to the use of standardized testing measures as criteria for improving educational outcomes, determining grade promotion, graduation, quality of schools, rewards or sanctions, ensuring equal educational opportunities, drawing in public support for schools, as well as many other highly attributable stakes and consequences (Heubert & Hauser, 1999). While there may be no current agreement or common consensus that a specific test is named a high-stakes test, it is recognized by many, that high-stakes are not identifiable characteristics of the test itself, but rather the effects of the intended and unintended consequences of the test scores (Goertz & Duffy, 2003; Heubert & Hauser, 1999).

The No Child Left Behind (NCLB) Act of 2001 legitimized the role of high-stakes testing through federal legislation and fueled the current debates over high-stakes testing. However, a salient factor contributing to the current reform debates may be that the term “high-stakes testing” does not appear in the hundreds of pages of the NCLB (2002) law (Johnson & Johnson, 2006). Instead this law states that responsibility in the form of standards and accountability is the main focus of the national educational goals. How we got from standards and accountability to mandated high-stakes tests across the

nation may be at the heart of understanding how high-stakes are defined by others (Linn, 2003).

Currently, high-stakes testing forms a core foundation which shapes American education policy. It is imperative, with the term *high-stakes* testing punctuating (Allington, 2003) almost every educational initiative or program, that a commonly held definition is at the core of effective and successful school reform. This study frames “high-stakes testing” as being those tests which critically impact programs, pedagogy, practice, curriculum, individual student achievement, teacher identity and resulting in high-stakes consequences within the educational setting.

In summary, high-stakes testing has informed the familiar experience of educational culture, thereby creating a situation in play where the familiar and common experience of schooling has created a situation in use where familiarity masks the complexity (Lowenberg Ball & Forzani, 2007) and use of the term high-stakes testing in schools operating within the standards reform movement. However, defining the federal view and epistemological roots of high-stakes testing is a positivist theoretical frame which looks to behavior and student knowledge as something to be measured – rapidly becoming the information kudzu of this century.

Situated Context of Study

Standards and accountability in the form of high-stakes testing in Tennessee are described as those state tests and other associated tests which hold accountability components in the form of sanctions and incentives which act as levers to improve the

overall quality and performance of education as a system. As the stakes attached to these tests increase, so do the real and perceived consequences for teacher pedagogy, practice, and teacher identity formation.

The stated purpose of the No Child Left Behind Act (2001) is to ensure that “all children will have a fair, equal, significant opportunity to receive a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state assessments” (NCLB, 2001). NCLB is based on three conceptual components: standards, assessment, and stakes (Cawthorn, 2007). Under No Child Left Behind (NCLB), schools and school districts are measured on whether students meet performance benchmarks in math, reading and attendance for grades 3-8, and math, English and graduation rates for high schools. Federal report card labels identify individual schools whether they have achieved Adequate Yearly Progress.

The No Child Left Behind Act (NCLB, 2002) builds upon the accountability provisions in the Improving America's Schools Act (IASA) of 1994 (IASA, 1994), which required each state to establish challenging content and performance standards and to implement assessments that measure students' performance against those standards (Goertz, 2001). The IASA defined adequate yearly progress (AYP):

[i]n a manner that results in continuous and substantial yearly improvement of each school and local education agency sufficient to achieve the goal of all children ... meeting the state's proficient and advanced levels of achievement; and is sufficiently rigorous to achieve the goal within an appropriate timeframe (Elmore & Rothman, 1999, p. 85).

Adequate Yearly Progress (AYP) status is calculated for the following student subgroups: White, Hispanic, African American, Native American, Asian/Pacific Islander, Economically Disadvantaged, Students with Disabilities and English Language Learners. If a school does not make AYP for two consecutive years, it is identified as a school needing improvement and targeted for school district intervention. If a school continues to not make AYP this school is then placed into corrective action. If a school does not attain AYP results after five years this school is placed under intensive sanctions and restructured. It is within this reform agenda that this survey research on the effects of high-stakes testing is conducted.

This study investigates the effects of high-stakes testing on teachers' reported pedagogy, practice and identity in the region of Eastern Tennessee. The state of Tennessee has a very complex school system that comprises 136 districts with 1,700 schools. The region known as the Eastern Field Services Division is comprised of 14 school districts including both county and city systems. High-stakes consequences were first attached to test instruments in Tennessee in 2000 (Amrein & Berliner, 2002). Tennessee conducts tests which may be viewed as high-stakes tests on several levels within the public school system. Although Tennessee states it has not yet set rewards and incentives attached to its high-stakes tests, the state does publically identify those schools which are deemed to be low performing schools. Punitive sanctions are leveled on low-performing schools as: "on notice", "probation" and finally, "state takeover". Additionally, while Tennessee does not regularly publish individual school test score

results, the state does employ a “well-publicized value-added assessment methodology” (Berry et al., 2003, p. 10). The Tennessee Value Added Assessment System (TVAAS) analyzes student test score data to construct teacher value-added measures of individual teaching effectiveness and estimates the effects of the collective work of teachers who work with a given student in a school year on overall score gains.

Currently, all states have adopted high-stakes testing policies and have mandated tests based on state standards as required by NCLB. Tennessee students in Grades 3-8 take the Tennessee Comprehensive Assessment Program (TCAP) Achievement Test each spring - student scores are reported to parents, teachers, administrators, and the public. The TCAP Achievement Test is a timed, multiple choice assessment that measures skills in Reading, Language Arts, Mathematics, Science and Social Studies. In one of the largest scale national surveys conducted on the issues of how teachers perceive the effects of high-stakes testing, Pedulla et al. (2003) identified Tennessee as a High-stakes/High-stakes state where “there are high stakes for districts, schools, and/or teachers *and* high stakes for students” (Abrams et al., 2003, p. 1). Although Tennessee contends that rewards or sanctions are not attached to high-stakes tests, the Pedulla et al. study found characteristics such as rewards and sanctions were in policy and practice at all levels within the state of Tennessee.

Over the last decade, the NCLB legislation has mandated standards, accountability and assessments which hold schools responsible to improve instruction and raise student proficiency scores on state standardized tests, thereby creating a

complex high-stakes educational environment (Linn, 2000). The mandates of NCLB identify schools that have missed a state standardized benchmark in the same category for two consecutive years. Tennessee elects to alert schools and districts that are at-risk of becoming a high priority school under NCLB, giving schools additional support and assistance from the state in order to avoid the NCLB high priority list. In Tennessee, schools that have missed one or more benchmarks for a period of one year are considered target schools.

Viewed as reliable and valid indicators of teaching and learning, state test results are mandated to be reported in scripted formats. However, these reported valid and reliable test scores are highly impacted by many local and state variables. As a result, these individual and aggregated test scores represent highly unstable intervening variables which may vary greatly in both standards of validity and reliability (Haney, 2000; Linn, 2003). Additionally, the consequential validity of the high-stakes test itself comes into question when viewed at the local level. Paris (2000), arguing that the consequential validity of a high-stakes test is considerably lower than policymakers and the public acknowledge, states that “even if assessments are rigorous in terms of psychometric validity and reliability, they may be low on consequential validity if the results do not help the stakeholders in positive ways” (p.5). While NCLB includes many aspects which are positive and supportive to systematic school reform, it is essential that teachers are included in the implementation and realistic setting of goals (Linn, 2003).

In Tennessee, teachers working within a high-stakes testing environment where test results are used for both state and NCLB accountability, have expressed concerns regarding the effects of the high-stakes testing environment. Nationally, students are tested more frequently than any other student group within the industrialized nations (Merrow, 2001). No issue in education appears to strike at the heart of real teaching and learning as does the issue of the effects of high-stakes testing (Hoffman et al., 2001).

Theoretical Framework and Conceptual Apparatus

Drawing upon the sociocultural theories of critical participation and action (Bourdieu, 1977), constructs of power and positioning (Foucault, 1997; Harre & van Lagenhove, 1999) and in critical conscientization of awareness and agency (Freire & Macedo, 1987; Holland, Lachicotte, Skinner, & Cain, 1998), I frame the development of and enacted teacher identity (Alsup, 2006; Britzman, 2003) within a complex social and culturally constructed perspective (Vygotsky, 1978; Wenger, 1998) of teachers' work in a complex and conflicted high-stakes testing environment.

Teachers working within the high-stakes testing environment are caught within the press of power, positioning and agency. Teachers working within the high-stakes testing environment find themselves within a powerful social and cultural construct where players are often found to be at odds with the structures of mandated reforms. The theoretical model which informs this research is influenced by a complex conceptualization of enacted teacher pedagogy, practice and identity (see Figure 1:01).

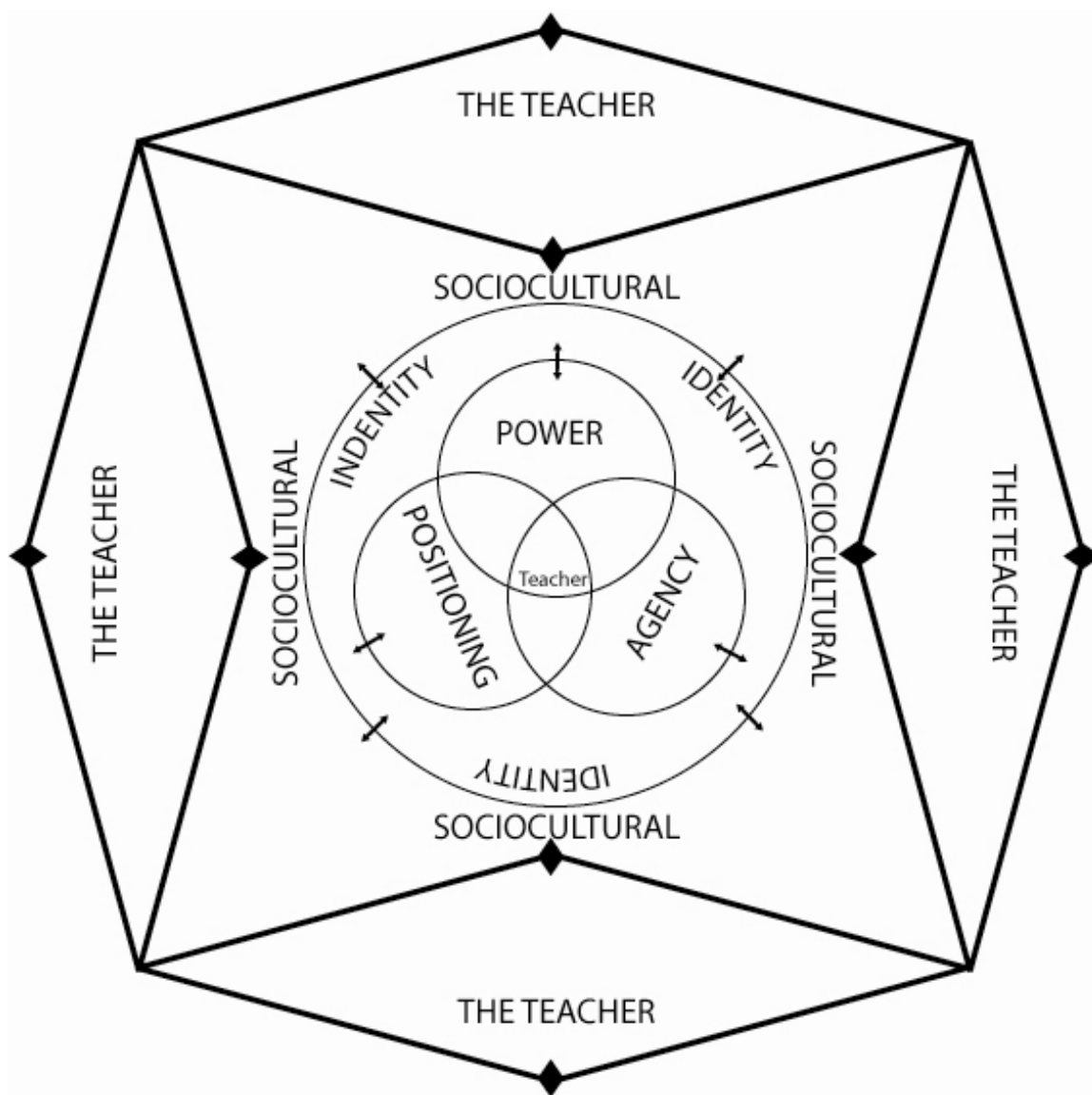


Figure 1:01: Theoretical framework and conceptual apparatus

This study is informed by sociocultural theories and conceptualizations, focusing on a purely non-positivist stance, which acknowledges the sociocultural nature and construction of human interaction. Bourdieu's (1977; 1996) theories of critical participation and action describe those cultural barriers which challenge and marginalize teachers while proffering a view to the marginalization of some and not others. Hence, the concepts of *habitus*, *capital* and *fields* work to shape, afford and challenge teacher participation and action within the high-stakes testing environment.

Acknowledging the active role of teachers within a system of power is viewed through Foucault's theories (1997) where the structure of power is acknowledged and named as it works to manifest itself and subtly and overtly exert itself upon an educational network. Positioning theories provide a perspective which places the teacher within a social, cultural and political metaphorical position framed by the high-stakes testing environment (Harre & van Lagenhove, 1999). Freireian pedagogy adheres to the role of the critical teacher, where critical participation is integral to action and/or agency, in response to oppression (Freire, 1995). Teachers situated within these roles are not passive bystanders but are active in self-authoring their own power, position and agency within the high-stakes testing environment; identity within this conceptual apparatus becomes a fluid, purpose driven characteristic of teachers' pedagogy and enacted practice (Holland et al., 1998).

Alsop (2006) describes a view of identity which "is holistic – inclusive of the intellectual, the corporeal, and the affective aspects of human selfhood" (p. 6). Teacher

identity recognized in this way is inclusive of all those who engage in and construct the experience within a temporal, responsive and reflective high-stakes teaching environment. Recognizing that teacher identity is formed within the complexity of relationships, Britzman (2003) proposes that "[e]nacted in every pedagogy are the tensions between knowing and being, thought and action, theory and practice, knowledge and experience, the technical and the existential, the objective and the subjective" (p. 26). Additionally to this perspective, Britzman (2003) illuminates that practice within this setting is a "paradox, an unanticipated social relation" (p. 3) where teachers shape their teacher identities in relation to the educational environment. Alsup and Britzman view identity as a centering of one's ideas, values, attitudes and beliefs, where the "moments of conflict and disjuncture are often the places where learning occurs" (p. 5). For Foucault (1972), this is the place where multi-layered histories, of self and others, present as highly contentious opportunities. He suggests that if one extorts for convenience, identity becomes consistent, ordered and de facto, losing all fluidity and interest in discovery and simply falling to linear, predetermined teleological views of pedagogy and practice. Foucault describes the fearful place of those teachers who hold fast to long-held, internalized narratives and beliefs of pedagogy and practice as one where power, positioning and agency is denied.

Vygotsky's (1962; 1978) social and cultural constructivist theories of learning and development describe the powerful role of social and semiotic mediation in shaping teacher pedagogy, practice, and identity. Vygotsky's theories highlight the seminal

importance of the social and cultural experience in the development of self and cognitive processes. Vygotsky believed that social mediation is at the heart of all learning, describing a zone of being (Zone of Proximal Development) which is highly negotiated within interactions of self and others. Here, the theories of Bourdieu align closely to the work of Vygotsky, recognizing that one's cultural and historical references play a situated and coexisting role in individual functioning and construction. From this perspective, teacher identity is mediated by interactions with others who are intertwined within the social cultural interaction of teacher's work. Pedagogy, practice and identity viewed from a Vygotskian theoretical frame realizes that social interaction is essential where experience is first processed on the social level and enacted on the individual level.

The theoretical framework and conceptual apparatus described, both influences and informs this quantitative survey research where pedagogy, practice and teacher identity is negotiated, mediated, and constructed within a dynamic social and cultural interaction (Vygotsky, 1962;1978; Wenger, 1998). Teacher pedagogy, practice and identity recognized as a multi-perspective construct when positioned within the Sociocultural theoretical paradigm of teachers' enacted practices brings the complex socially constructed, mediated nature of participation to the forefront of this study.

Illuminating experiences of the respondents within these school communities of practice (Lave & Wenger, 1991) through a self reported survey methodology provides teachers the opportunity to construct or author the telling of their experience within the high-stakes testing context. Conceptualizing that "multiple realities exist that are

inherently unique because they are constructed by individuals who experience their world from their own vantage points” (Hatch, 2002, p. 15) is recognized within the instrumentation of the research. Additionally, the Sociocultural paradigm offers a theoretical foundation from which the self-reporting, anonymous quality of a survey methodology works naturally from, allowing for those moments of *discontinuity*, *heterogeneity* and *variability* that are inherent to teaching and learning (Foucault, 1972) to be ‘voiced’ and recognized.

The contextual nature of teacher pedagogy, practice and identity is produced within school communities where respondents are actively engaged within the high-stakes testing environment. From this theoretical and conceptualized perspective, the sociocultural paradigm recognizes social and cultural influences and interaction on the construction of teacher identity within a setting which includes power, positioning and agency.

Purpose of the Study

The purpose of this quantitative survey research was to compare and examine the relationship of the consequential effects of high-stakes testing on teachers’ pedagogy, practice, and identity. Critical to developing an understanding of teachers’ perceptions and experiences, in terms of high-stakes testing in current educational settings, is documenting the “voice” of the respondents.

Specifically, this study places special emphasis on the perceptions and experiences of teachers as they mediate and as they negotiate the high-stakes testing

environment in Tennessee. The practical knowledge of teachers in Eastern Tennessee schools was documented in terms of their pedagogical, professional, and perspective response to their experiences within the current high-stakes testing environment. Factors which mediate these experiences are posed using a semi-structured and open-ended survey methodology. How teachers enact teacher identity and agency as they hold to the pedagogy that informs their practice grounds this research. The focus and purpose of this quantitative survey research is to document the experience of teachers who are engaged in state mandated high-stakes testing. By documenting the voices of these teachers, this study gives voice and clarity to the experiences of those teachers who are implementing the high-stakes tests.

Research Questions

The primary purpose of these research questions is to describe and explore the pressing questions of teachers' perceptions within the high-stakes testing environment.

The central research questions of this study are:

3. What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?
4. What are the consequential effects of high-stakes testing in relation to teachers' work and identity?

Sub-questions based upon specific variables define the domains of inquiry within the survey instrument:

1. Perceptions:
 - a. What are the perceptions of teachers in relation to the effects of high-stakes tests?
 - b. Do teacher perceptions of high-stakes tests differ by independent variables (setting, grade, experience and school performance)?
2. Actions:
 - a. What actions relating to pedagogy and practice are teachers taking (preparation, time and mode)?
 - b. Do actions of preparation, time and mode differ by independent variables (setting, grade, experience and school performance)?

Significance of the Study

Previous research studies have raised a number of issues focusing on broad questions seminal to situated issues and concerns, such as, high-stakes testing in relation to instruction (Corbett & Wilson, 1991), curricular control (Au, 2007), validity of high-stakes tests (Amrein & Berliner, 2002a), and the changing roles of teachers (Barksdale-Ladd & Thomas, 2000); however, less has been documented on teachers' perspectives and experiences as a result of high-stakes testing. The prevailing assumptions that high-stakes testing is both an appropriate measure of student achievement and learning continues to move forward without representative data to support its mandate (Riddle Buly & Valencia, 2002).

Currently, standardized tests have come under a barrage of criticism; misuse of test results has fueled the arguments of heavy reliance on test results where high-stakes consequences are the outcomes (Nichols & Berliner, 2005). Mandated testing within the

reform agenda has become a socially constructed symbol to represent that something is being done (Madaus, 1988b). While many scholars and researchers currently debate the definition and significance of high-stakes testing, the utilization of large-scale high-stakes tests has become common in American schools (Heubert & Hauser, 1999). Despite reported negative consequences and misuse, standardized tests, when appropriately utilized, can provide helpful information and identify trends and patterns for larger instructional and program planning (Cimbricz, 2002).

Implementation of mandated high-stakes tests delineates a reform movement which is complex and ideologically representative of any large historical social change where government is naming policy which directly impacts contemporary educational ideologies, policies, and structures (Au, 2007). Currently, little is currently understood regarding the ways in which teacher identity is influenced and impacted by the high-stakes testing environment (Lasky, 2005). As well, the demands and pressures on teachers working in the high-stakes grades and subject areas so deeply impacts the educational setting that it is important that we understand both the limitations and the potential of the consequential effects of high-stakes testing (Pedulla et al., 2003; Stecher & Barron, 2001; Taylor, Shepard, Kinner, & Rosenthal, 2003).

This study utilized a previously validated national survey instrument taken from the Pedulla et al. (2003) study. The Pedulla et al. research was conducted during the first years of the NCLB mandates; now, with six years of reform implementation following this previous research, it is prudent to undertake a study examining the effects of high-

stakes testing. While this survey research does in no manner seek to replicate the previous Pedulla et al. study, two recommendations for future research from Pedulla et al. study are recognized:

1. further research needs to include teacher voice
2. further research and examination of the effects of high-stakes testing at the local or state level is needed.

In response to these recommendations from the Pedulla et al. study, two important features of this current research are noted. First, the need for this quantitative survey research to contribute important dimensions of teacher voice to the developing understanding of the effects of high-stakes testing on teacher pedagogy, practice, and identity. Secondly, this study seeks to contribute a teacher voice which has been largely absent from the local and state level reform conversations as policymakers have regarded these data sources as “too personal, idiosyncratic or soft” (Goodson, 1993, p. 10).

Support for these research purposes is reflected in a statement by Arnold Shore, executive director of the National Board on Educational Testing and Public Policy, stating that “[i]n the public debate, in the public conversation, the voices of those who are implementing testing and accountability policies are either under heard or not heard much at all” (Shore, In Olson, 2002, p. 2). Further, recognizing the scant representation and importance of teacher voice in educational research and policy-making, Clandinin and Connelly (1985; 1995) continue to call for those who question the efficacy of the current

reform movement in the United States to include the teacher's voice regarding their personal practical knowledge within the context of the educational setting.

Critical to understanding how teachers have responded to the high-stakes testing mandates and how these measures have effected their teacher pedagogy, practice and teacher identity as they work within these accountability systems is examining these effects from the teacher perspective. Most importantly, this study promises to contribute and extend the current research and foster a greater understanding of the relationships among teacher perceptions and experiences within the high-stakes testing environment. Findings may suggest ways and means to influence and contribute to future policy and practice.

Assumptions and Researcher Ideological Stance

Several assumptions underlie this study. First, the researcher assumes the respondents of this study are a representative sample of Eastern Tennessee teachers. Second, it is recognized that there are several tangible and intangible variables that can impact the study such as: school leadership, professional development, teacher experience, individual and group resilience (Allington, 2002; Fullan, 2006; Wasley, 1991). Third, it is assumed that the self-reported responses are sufficiently accurate and free of error. Fourth, it is assumed that the survey instrument accurately measures teacher's perceptions of the effects of high-stakes testing on teacher identity, pedagogy, and practice.

Additionally, it is essential to describe the epistemology of the research methodology which undergirds this study. “Underlying (any research methodology choice) is a view of knowledge that influences the development of research questions, the data sources, the data-collection procedures, and types of analysis” (Dressman & McCarthy, 2004, p. 340). As a researcher, beyond any assumptions, it is essential that my ideological beliefs are stated, as well as making clear my epistemological foundations, where influential characteristics may stand as intended or unintended variables underlying the methodology and viewed as bias within the study.

The primary ontological characteristics of survey methodology describe the reality as it is documented by the respondents within the study (Nardi, 2006a). This ontology creates an authentic foundation for the survey study as it is designed. Additionally, the epistemological consideration of the relationship of the researcher to the study itself is one of importance. Cunningham and Fitzgerald (1996) pointed out that developing a deep understanding of epistemological underpinnings “can bring to light presuppositions and assumptions that might otherwise go unquestioned ... further, clarifying one’s own epistemological stance as a researcher and/or teacher can lead to a self awareness and understanding” (p.36) gives further strength the study as a whole. This is an important challenge in my role, as researcher, further highlighting the importance of a clear demonstration and awareness of my own subjectivities. This study is framed by “[a] sensitive awareness of the methodological literature about self in conducting inquiry, interpreting data, and constructing the final narrative helps” (Marshall & Rossman, 1995,

p. 19) to ensure the generalizability of the findings. The importance of documenting and describing the experience from the respondent perspective is central to the generalizability of this quantitative survey research.

I draw upon sociocultural theories to analyze and document the effects of high-stakes testing on teacher pedagogy, practice, and identity. Analysis of the data is framed by Stone's (2002) epistemological stance of numbers as metaphors. Dispensing here with the assumption that high-stakes tests count what is "right and best", the focus of this study recognizes that we must be just as interested in the *discontinuity*, the *heterogeneity*, the *variability* that is inherent to teaching and learning (Foucault, 1972) within the high-stakes testing environment. Within the scope of this study, as the researcher, I have documented teacher voices within the methods and characteristics of survey methodology as related to their experiences. Choosing to view the empirical data within a metaphorical stance presents a seemingly provocative paradox where this study stands as a kind of "metaphorical praxis" between theory and practice, intent and consequence, education and schools. As a researcher, I am interested in what is essential and central to these respondents within the framing theoretical underpinnings of a sociocultural perspective where power, positioning, and agency is at the center of enacted teacher identity within the current high-stakes testing environment.

Like all researchers, I bring my personal and professional knowledge, beliefs, and values to this study. My many years of experience in elementary, middle, high school and university level teaching contribute to my role as researcher. I am aware that my teacher

backstory and years of professional experience represent a foundation of values, beliefs and experiences relating to high-stakes testing practices. Furthermore, my professional experiences as teacher, curriculum developer, coordinator and project leader have defined, for me, a model of orientation, implementation, and reform practices which have “best practice” characteristics. These factors have influenced all aspects of this study – research topic, formulation of the primary research questions, and way in which I have conceptualized and designed this study.

As the researcher, I began this study stating my subjectivities and biases, being open to the multiple realities and experiences that may have existed, and subsequently applied research methods which worked in a quantitative way to document what was important to these respondents, and finally, document knowledge, resulting from the lived experiences of teachers working in the high-stakes testing environment. I remain mindful that, “when you are standing within the circle of logic created by the assumptions of your paradigm, the positions taken by those working in other paradigms simply do not make sense” (Hatch, 2002, p. 19).

Importantly, I am cognizant of the complex challenge of documenting and describing others’ experiences. The very nature of survey methodology is one which depicts and documents the told reality of the respondents. Stake (2000) asks us to be aware that researchers sometimes “pass along to readers some of their personal meaning of events and relationships – and fail to pass along others” (p. 442). The self-report method of survey methodology allowed the researcher to capture the told perspective of

each respondent in this study. With this caution exposed, I have continued to be aware and make every effort to maintain the stance of researcher within this study.

As a result, several intentional research strategies were utilized to mitigate the effects of researcher bias. First, the survey instrument was sent to a representative sample of teachers who did not have face-to-face contact with the researcher. This strategy served to ensure a greater quality of respondent response by removing interviewer bias (Mitchell & Jolley, 2007). The anonymous respondent format allowed for less social desirability bias, where respondents may have responded how they perceived the interviewer may have wanted them to (Mitchell & Jolley, 2007). Additionally, the survey instrument utilized in this study had been subjected to rigorous validation studies to ensure that the instrument is measuring what was intended (Pedulla et al., 2003). Finally, combining rigorous, transparent data collection and analysis procedures, as well as being critically aware of my stance as researcher, served to minimize bias in this study.

Limitations of the Study

Limitations of the study are viewed as constraints imposed by the data collection process. The quantitative findings of the study are:

- limited to the responses from a sample of Eastern Tennessee teachers
- limited to the data resulting from this survey methodology
- limited by the ‘gatekeeper’ status of the principal
- limited by the voluntary, self-report nature of this survey design
- limited to non-causal findings

Delimitations of the Study

Delimitations of the study determine the context or parameters of the study such as: population sampled, selection criteria, and demographic data included in the data analysis. This study:

- does not attempt to survey outside of the sample parameters
- does not attempt to generalize outside of the generalizable population
- does not determine or evaluate the preparation or training of respondents
- does not attempt to address the difference in student populations and academic abilities

Definition of Terms

Accountability: Accountability measures and indicators are established in the form of standards, goals, and performance.

Adequate Yearly Progress (AYP): State measure of yearly progress based on mandated state academic standards. AYP is the minimum level of improvement that states, school districts, and schools must achieve each year. Schools that do not meet the achievement standards for two consecutive years are deemed high priority.

Agency: Human agency is contingent upon the possibility that a person “could have acted otherwise” (Giddens, 1979, p. 54) within a setting that is situated in a time-space relationship inherent within an agentic social interaction.

Capitals: Conceptualized as: *cultural, social, economic* and *symbolic*; these are the key factors that define position and possibility.

Fields: A space of social and cultural relationships which impact and influence participation and action.

Habitus: A set of internal prior experiences, structures and schemes of thought and action which work to regulate and shape participation and action.

High-stakes Tests: Those tests which have important consequences for the test-taker and may in turn critically impact programs, curriculum, individual student achievement, and teacher's professional lives resulting in high-stakes consequences.

Identity: Identity existing within a political, social, historical, cultural, and ideological endeavor, is a multi-faceted construct that values specific ways of constructing identity, while silencing others. Power, positioning and agency viewed within a sociocultural lens is an embedded element of identity formation, where identity as a fluid, shifting construct is recognized differently by others. Identity is unstable as it exists in relation and is situated socially.

National Assessment of Educational Progress (NAEP): To remain eligible for federal funding under the No Child Left Behind Act, states are required to participate in NAEP reading and math assessment every two years at the 4th and 8th grade levels. If academic achievement does not change after high-stakes are attached to a state test or if achievement decreased, the effectiveness of the high-stakes policy as a means of improving student performance is called into question. State NAEP 4th and 8th grade scores allow the public to judge individual states'

academic progress and compare state achievement scores against one another.

No Child Left Behind Act (NCLB): The NCLB Act was signed into law January 8, 2002. The purpose of this Act is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on State academic achievement standards and state academic assessments (<http://www.ed.gov/nclb>).

Pedagogy: Inscribing into practice strategies of instruction or style of instruction which is directly influenced and impacted by theory, beliefs, knowledge, values, mandated policies and teaching experiences (Pahl & Rowsell, 2005). Pedagogy refers to the ‘what’ of teachers practice; comprised of the curricular choices teachers bring to the teaching and learning act.

Positioning: “[S]ymbolically mediated interactions between people, both from their own individual standpoints and as representatives or even exemplars for groups” (Harre & van Lagenhove, 1999, p. 1).

Power: “When I speak of relations of power, I mean that in human relationships ... power is always present: I mean a relationship in which one person tries to control the conduct of the other... these power relations are mobile, they can be modified, they are not fixed once and for all ... (they are) thus mobile reversible, and unstable”(Foucault, 1997, p. 292).

Practice: Referring to the act and/or the enacting of ways of teaching and working as a teacher. Practice refers to the ‘how’ of teachers work.

Reform Movement: Refers to the current rapid, mandated, policies and teaching practices in place across America, rather than a kind of social movement or change which is gradually implemented through local needs and assessments.

Standards Movement: A Nation at Risk emphasized the development of standards for student performance in these areas:

- **Academic standards** describe what students should know and be able to do in the core academic subjects at each grade level.
- **Content standards** describe basic agreement about the body of education knowledge that all students should know.
- **Performance standards** describe what level of performance is good enough for students to be identified as advanced, proficient, below basic, or by some other performance level.

Tennessee Comprehensive Assessment Program (TCAP): A criterion-referenced test which is used for both state and NCLB accountability. The TCAP is a mandated K-12 state test which currently includes the Achievement Test (grades 3-8), the Writing Test, the Competency Test, the Gateway Tests and the End of Course Tests. The acronym TCAP is associated with high-stakes testing.

Voice: “[T]he concept of voice spans literal, metaphorical, and political terrains” (Britzman, 2003, p. 66) where teachers exercise and act upon their valued *right* to voice and have represented and described their experiences, beliefs and perceptions.

Organization of the Study

This study makes use of survey methodology to guide a research inquiry documenting the dynamic voices of teachers currently engaged in school communities, using quantitative data collection procedures to document an empirical and highly human perspective surrounding the study question. I am interested in what is essential and central to these respondents within the framing theoretical underpinnings of a sociocultural view. This study seeks to identify and describe effects of high-stakes testing through the documentation of the voices of teachers. Choosing to utilize a survey methodology for this study offers an opportunity for teachers' voices to be documented and heard within a safe and anonymous response forum. Specifically, structured and semi-structured survey methodology affords an inquiry method where several sub-questions may be asked in relation to the central study question to further describe and document the characteristics of the primary research question (Baumann & Bason, 2004; Dillman, 2007; Nardi, 2006). Finally, I see this research question as a highly contemporary issue and social concern, which may potentially move the dialogue beyond the educational context. I have explored this question to document the voices of the teachers asking: What are the consequential effects of high-stakes testing on teacher pedagogy, practice, and identity?

In overview, Chapter 1 presents an introduction to the current high-stakes testing environment, focusing of the study and providing a background to the problem, purpose of the study, describing the historical context, operationalized researcher's definition of

high-stakes testing, stating the research questions and theoretical framework, limitations, delimitations, and study definitions. Chapter 2 presents a substantive review of the comprehensive literature providing a foundation for the research. Chapter 3 discusses the decision to conduct a quantitative survey study in relation to the research questions. Additionally, Chapter 3 focuses on the research design, instrumentation, sampling frame, sample, data sources, methods of data collection and data analysis. Study findings and results are presented in Chapter 4. The conclusions and implications of the results and recommendations for future research agendas are discussed in Chapter 5.

CHAPTER 2

REVIEW OF THE LITERATURE

Chapter Introduction

Examining the effects of high-stakes testing on teacher pedagogy, practice, and identity is essential to the understanding of how teachers view their work within the current high-stakes testing environment. The following chapter focuses on three broad areas of empirical literature which have influenced the development of this study: teacher identity, literature examining the effects of high-stakes testing from diverse lenses and literature which has utilized quantitative methodology. These areas of research are intended to define the focus and position the reader within a review of the literature which is relevant to the study of the effects of high-stakes testing on teacher pedagogy, practice, and identity.

Central to this review is the concept of teacher identity as recognized within teachers' work in the high-stakes testing environment. The first section of this chapter looks closely at the how educational researchers have recognized the concept of teacher identity to study educational pedagogy and practice. Section two presents an overview of the proponent and opponent research and examines a representation of the methodological diversity within the current research studying the effects of high-stakes testing on teacher pedagogy and practice. The third section reviews educational research which has specifically utilized quantitative survey methodology to examine the effects of high-stakes testing. The research presented represents a portrait of the seminal research in

the field. Research which held to specific characteristics of rigor and research standards of quality (Howe & Eisenhart, 1990; Shanahan, 2000; Wortman, 1994) were studied and examined, thereby justifying inclusion in this literature review.

Using standard search terms produced a large amount of scholarly writing on this subject. Research strategies employed to identify the relevant literature related to high stakes testing included ongoing computer searches conducted through multiple educational databases (Wilson Education Abstracts and ERIC), *Google Scholar*, and *Dissertation Abstracts International*. A consistent and thorough searching of individual reference bibliographies from selected books, articles, published and unpublished dissertations, research reports and papers was thoroughly conducted.

Further focusing of this extensive search of the literature, the researcher chose to review studies which presented a diversity of methodologies and which maintained standards of research quality; this parameter reduced the works to a more selective body of qualitative and quantitative research. Extensive methodological and analytical diversity was considered in relation to the central research question. Each research study was considered for its ability to contribute to and support the central research inquiry of this study viewed through the lens of Howe and Eisenhart's (1990) standards of research quality:

1. Rigorous data collection and analysis methods applied.
2. Explicit contributions from existing theories, exemplary research, other relevant literature from the field and researcher subjectivity (bias).
3. Alignment between research questions, data collection and data analysis.

4. Overall validity established.
5. External (purposeful) and internal (ethical) value constraints must be addressed.

The search of the literature generated a vast amount of educational research on the topic of high-stakes testing, teacher attitudes and beliefs of high-stakes testing, teacher identity, educational reform, etc. This review of the literature examines the selected quality and relevant research from the field which has explored teacher identity and high-stakes testing in relation to teacher pedagogy, practice and identity.

Teacher Identity Framed

The purpose of this first section of the review of literature is to better understand how teacher identity formation relates to the effects of high-stakes testing and teacher pedagogy, practice and identity. This section of the review of literature is grounded in sociocultural theories of identity construction, specifically critical theory and constructs of power, positioning and agency. Educational theorists and researchers have studied teacher pedagogy and practice to offer valuable insight to the understanding of teacher identity. Researchers have recognized teacher identity within the complex social and cultural dimensions of educational settings where teachers' ways of positioning themselves accounts for the essential qualities of teacher identity. These enacted qualities are "deeply constitutive of people's social and psychological being" (Shotter, 1989, p. 142) where the context of teacher's work plays a significant role in the embodiment of their teacher identity.

How teacher identity influences and impacts pedagogy and practice is an essential, yet somewhat intangible construct. In a recent study examining the effects of teacher identity, Hammerness, Darling-Hammond, Bransford, Berliner, Cochran-Smith, McDonald, and Zeichner (2005) explained that teacher identity can “shape their dispositions, where they place their effort, whether and how they seek out professional development opportunities, and what obligations they see as intrinsic to their role” (p. 384). Cochran-Smith (2003) described the complexity of teacher identity as “the ways we stand, the ways we see, and the lenses we see through” (p. 2). Similarly, Wenger (1998) distinguished a “profound connection between identity and practice” (p. 149) describing “the experience of identity in practice is a way of being in the world” (p. 151). Gee’s (2001) theories of identity and group membership recognize identity as what names or is essential to our being, how we see our sameness or difference within d/Discourses. Additionally, the Vygotskian (1962; 1978) perspective sees the powerful role of social and semiotic mediation in shaping identity, where identity is mediated by interactions with others who are intertwined within the social cultural interaction. Identity framed in these theoretical and conceptual perspectives is recognized as a constructed social practice evolving within a teacher’s pedagogical and practical experience.

Current educational research stemming from this paradigmatic thinking has “emerged as a heterodox diaspora, ideas and practices of which still coexist and intermingle with the didactic, behavioral science, paradigm” (Hamilton & McWilliam, 2001, p. 18) of earlier pedagogical research. Proposing a pedagogy for identity

development, Danielewicz (2001) explains the continual process of how identities arise with the action and interaction of sociocultural relations and shaped by the interplay of internal and external discourse. Recognizing teacher's "powerful, yet complicated and uneven histories" (p. 1) Danielewicz puts forth an argument which places teachers directly within the powerful, complicated and uneven mandated high-stakes testing environment and speaks of awareness, agency and possibility. Teacher identity from this perspective places teachers squarely at the helm of who they are and who they choose to be within the complex high-stakes testing environment.

Identity has been recognized as the "presentation of self in a matrix of social relationships – a pattern of social assertion that significant others recognize and come to respect" (Davidson, 1996, p. 2). From this perspective, Brewer and Gardner (1996) believe that connectedness and belonging are central to the concept of identity within the sociocultural setting. Therefore, teacher identity is recognized in terms of the reflective, changing pedagogic and practice enacted within the socially and culturally responsive school environment. Educational researchers have used identity to study the relationship between teacher pedagogy and practice and the ways teachers define themselves through their perceptions of self, relations with others, and with the greater educational system.

Crocco and Costigan (2007), in their recent study of New York teachers, state that teachers working within the high-stakes testing environment reported their "professional identity thwarted, creativity and autonomy undermined, and ability to forge relationships with students diminished" (p. 512) as a result of their work within the high-stakes testing

environment. Examining what becomes of teacher identity within a high-stakes setting, Salinas (2006) contends that teachers, asserting their professional identity, selectively complied with testing mandates to strategically negotiate and mediate the high-stakes testing environment. Teachers working within the high-stakes testing environment are challenged at many levels; understanding the complex nature of teacher identity formation may provide a greater insight to the social and cultural influences which impact teacher's work within the high-stakes testing environment.

The following section of the review of the literature examines six seminal research studies grounded in sociocultural theories of identity construction and constructs of power, positioning and agency. These researchers, having recognized the complexity of teacher identity, have explored the concept as a highly fluid, multi-layered, social and cultural construct within their research. The studies selected for review inform and frame teacher identity in relation to pedagogy and practice.

Identity and Process

In a qualitative case study Assaf (2005) investigated one preservice teacher's identity formation within a reading specialization program. Drawing on the theories of language and learning this study was informed by the work of Bakhtin (1981) and Gee (2001). Following an interpretive perspective the researcher asked the following research questions:

1. What is the nature of one preservice teacher's discourse in a reading specialization program?

2. How does identity influence the choices one makes and instructional practices one uses as a teacher?
3. What does one's discourse in a reading specialization program reveal about learning to become a teacher?

The teacher (Adrianna) in this single case study was an immigrant 25-year old preservice teacher who found herself enrolled in a reading specialization program in an effort to improve her grade-point average and move on to Graduate School in another field. However, after participating in the program for a short term, Adrianna decided to pursue teaching as her career. The researcher entered this study curious as to “how this transformation occurred and how the reading specialization program may have influenced her identities as a teacher” (Assaf, 2005, p.205).

Data were collected over an 18-month period from multiple data sources: observations and class meetings, archival documents, online discussions, instant messages, face-to-face interviews, journals, web portfolios, and the researcher's reflective journal. As part of her participation in the reading specialization program, Adrianna participated in computer-mediated discussions (CMD). These online discussions were included in the data set and used to explore dialogue and identity as it developed in the course of her participation in the reading education program. Data analysis, ongoing and designed in four stages, utilized the constant comparative method and discourse analysis. Stage one was the initial stage of coding the data into categories such as “phrases that illustrated tensions between internally persuasive and authoritative discourses” (Assaf, 2005, p. 207). Stage two was face-to-face interviews with Adrianna,

acting as a member check in verifying the initial findings from stage one analysis. In the third stage, interview data were compared with initial themes identified. In this stage, using a computer assisted qualitative software program (Nudist), data were reanalyzed to ensure the reliability of coding and categories. The fourth stage utilized Gee's (1999) methods of discourse analysis to explore the social, cultural, and situated discourse of Adrianna's experience.

Findings from the analysis report how Adrianna used language to mediate her "lived experiences, assumptions, and deeply held beliefs that shaped her identities as a teacher" (p. 202). Addressing the initial study questions, Assaf explores Adrianna's professional development and identity construction as revealed in the data. Assaf concluded that Adrianna's identity determined the pedagogic choices and practices she enacted as student teacher. Adrianna's decisions for teaching literacy were highly influenced from her own experiences as a child. Further, Assaf states that recognition and membership within the reading specialization program influenced and shaped Adrianna's self image as a reading teacher and sustained her commitment to a child centered pedagogy and practice.

As a new teacher, Adrianna grappled with the validity of high-stakes testing. She anticipated a loss of agency as a teacher and encouraged her peers to stand up and "teach against the grain" (p. 214) expressing a level of anger and advocating "revolution" (p. 214). Adrianna's strong sense of who she is and who she was not willing to become was evident in her discourse, "I refuse to believe what we are doing with kids [high-stakes

testing] is right. We must change it. Let's start with taking personal responsibility ... then maybe a revolution" (p. 214). For Adrianna, this was her agentive moment where hope and possibility presented in a call to her peers to "reinvent schools and change current inequities" (p. 214) within the high-stakes testing environment.

In conclusion, Assaf recognizes how these discourses represented Adrianna's teacher identity formation and how she was coming to see herself as a teacher within the complex educational setting. Identifying a current void in the literature, Assaf expresses the need for additional research to "investigate learning to teach as an identity-forming process" (p. 234). Additionally, Assaf states that further studies in the area of teacher identity formation can influence beginning teacher's understanding of the complex process which "influences their instructional decisions, literacy instruction, and commitment to children" (p. 234). Assaf's concluding emphasis for further research calls for a greater understanding of how beginning teachers develop in specific ways and what characteristics contribute to their learning and identity formation.

Identity and Commitment

The challenge of sustaining commitment amidst the influences of reform, standards, and identity formation is explored in a qualitative international study conducted by Day, Elliot, and Kington (2005). This study asserts that identity formation is:

a crucial element in the way teachers construe and construct the nature of their work; that commitment is a necessary element of professionalism; that motivation, self-efficacy, job satisfaction and commitment are closely linked with identity; and that teacher identities are the result of an

interaction between personal experiences, and the social, cultural and institutional environment in which they function on a daily basis. (p. 566)

These researchers, interested in looking at the complex phenomena of teacher identity and commitment, asked: What were the factors influencing teachers' challenges to sustaining commitment? Moreover, Day and his colleagues were curious if experienced teachers held characteristics of identity associated with hope and enthusiasm which influenced their abilities to sustain commitment to their work within a reform agenda.

Study respondents included 20 experienced teachers, ages 45 to 55, having between 25 to 35 years of experience, representing inner city, urban and rural contexts from Australia and England. Teachers were selected on an opportunistic basis, drawing from those who were involved in ongoing professional development at the researcher's universities. This sample comprised 6 primary teachers and 6 secondary teachers from Australia, as well as 4 primary teachers, 3 secondary teachers, 1 head teacher, and 2 district administrators. Data sources included in-depth interviews, field notes, and archival documents. Semi-structured interviews were conducted asking:

1. How do teachers themselves characterize those who are committed and those not so committed?
2. What has shaped teachers' levels of commitment and what sustains/diminishes this?
3. How do teachers characterize the changes in their levels of commitment across time?

Each interview was electronically recorded, transcribed and archived. Data analysis was performed utilizing an inductive approach where “analysis and interpretation of all data formed an ongoing part of the data-gathering process” (p. 568) to establish emerging themes and patterns.

Findings from the data were then reported and categorized by three analytical themes:

1. characterizing commitment
2. changes across time
3. factors which sustain and diminish commitment

Teachers expressed central themes associated with commitment as being “enthusiasm, belief in an ideal (vision), hard work, a sense of social justice, and an awareness of the need to attend to their own continuing development, and a recognition of priorities” (p. 570). The second broad theme, of change across time, was significantly associated with the theme of commitment. Teachers revealed that their commitment increased over time, fostered by their greater experience and ability to reflect. Teachers expressed they were further supported by their developing identity as a teacher. Additionally, teachers reported that with experience came the realization that some things were just out of their control.

Factors which sustained or diminished commitment, identified as the third theme, proved to be a unifying expression from both groups where both English teachers and Australian teachers expressed that “effective teaching was a far more complex mix of internal and external factors” (p. 571). Teachers articulated these factors as: level of

appreciation, institutional support, capacity to reflect, and the desire to be a learner.

Overall, the moderating effects of these factors served as an internal negotiator in their development of their teaching self.

Day and his colleagues suggest that commitment may be more wholly understood as a nested phenomena where there is a relatively permanent set of values based upon a person's beliefs, images of self, role and identity at the center. This study recognizes that teachers working within educational settings during times of change experience highly unstable professional identities and at times "mobilize 'occasional identities' in response" (p.575) to change. These researchers suggest that understanding teacher commitment is a key to a deeper recognition of teacher identity where values are subject to change when impacted by reform agendas which challenge one's beliefs, images of self, and identity. Day et al. conclude that during times of change teachers require additional support and networks to sustain their "collective identity and commitment" (p. 575) within the imposed reform agendas. This research affirms the complex nature of teacher identity within educational reform settings.

Identity and the Given and the Possible

Focusing on the lived experience of preservice teachers, Britzman presents a critical ethnographic narrative which focuses on two student teachers, Jamie Owl and Jack August. Positioning this study within the complex anxieties and conflicts of the educational landscape, Britzman offers a rich and contextual portrayal of teacher identity within educational structures. Britzman (2003) argues:

to consider what it is that structures the discursive practices of those learning to teach requires a double consciousness of persons, structures, and of the discourses that join them, and an acknowledgement of how the inadequacies of the present structure work through the practices of the newcomers (p. 221).

Presenting identity formation of teachers as a highly paradoxical occurrence, Britzman describes how preservice teachers are “inducted” into the teaching profession by more seasoned and experienced teachers. These teachers often presented a kind of “do as I say not what I have done” hindsight to novice preservice teachers.

This study looked closely at the contradictory experiences of learning to teach in a high school setting and how these experiences shape teacher identity. Anchored in lived teaching experiences, Britzman theorized and framed this study in the situated nature of local knowledge and how cultural myths and beliefs about teaching “are discursively produced and lived, and how the conditions of learning to teach inscribe the subjectivities, voices, and practices of its subjects” (p. 33). Britzman articulates two questions to instigate this inquiry:

1. What is it like to learn to teach?
2. What does it mean to those involved?

Articulating the influence of socialization on teacher identity, Britzman explains that the view of becoming she advocates “is not limited to what happens to persons. Instead, [her] concern is in understanding what they make happen because of what happens to them and what it is that structures their practices” (p. 70). Here Britzman recognizes that within a highly conflicted power setting, teachers have the power to

mobilize through these contentious realities and shape their pedagogy and practice as they see fit. In this way, teachers are able to mobilize and reshape their teacher identity beyond a binary view of becoming.

Following an ethnographic, narrative research methodology, Britzman describes a discerning portrayal of these two student teachers. She includes an insightful and essentialized portrayal of teacher identity formation, conveying the influence of personal biography. While the respondents did not speak of race, class, or gender in a conscious way, Britzman identifies the inherent and influential nature of these categories and carefully describes the unique and discreet characteristics of each respondent in terms of race, class, and gender. Indeed, Britzman further asserts that “teachers are also supposed to “shed” their own social casings and personal preferences” (p. 234) as they take on a normalized teacher identity. The student teachers in this study expressed feeling that there was “no way out of the reproductive cycle” (p. 236) where expectations and the dynamics of cultural reproduction set them in collision with authoritative pedagogy or to maintain the status quo. However, this portrayal of these teachers as wholly impressionable and without a strong sense of self, leaves one to question how they negotiated and mediated the tensions of such a suppressed identity throughout the practicum experience. Britzman discusses the influence of individualism on pedagogy and practice choices where student teachers were “caught up in the tension between subjectivity and objectivity, and thus could not interpret the pedagogical issues this dualism positions” (p. 105). Student teachers, in this study, were faced with a kind of

compliance to the “reproduction of school structure through pedagogy and the suppression of any differences that can move one toward a dialogic understanding of pedagogy and the self” (p. 237).

Findings from this study indicate three myths that surround teachers’ identity and practice: everything depends on the teachers, the teacher is the expert, and teachers are self-made. Further to this, Britzman contends that each of these myths

authorizes a discourse of power, knowledge, and the self that works to promote the impossible desire of assuming the self to be capable of embodying a noncontradictory subjectivity and capable of asserting a form of control that depends upon the individual’s unambivalent acceptance of authoritative discourse (p.223).

The myth of “everything depends on the teacher” assumes that the teacher is in control and always operating on high alert to avoid any deviation from the planned. “The problem is that within this push to control learning, the student teacher must devalue her or his own power to explore with students the dangerous territory of the unknown” (p. 224).

The myth of the teacher as expert exemplifies a primary source of conflict and anxiety for student teachers. Students enter into an experience where they are expected to be experts rather than students engaged in a practicum experience to refine and expand their emerging profession identity and practice. Britzman identifies this socialized view as contrary to the notion of practicum and experience building where the student teaching experience is not seen as “an intellectual, emotional, and aesthetic challenge, but as a function of accumulating classroom experience” (p. 229). Further to this, Britzman

exposes the duality of this experience where students are “being educated as a student while educating others as a teacher” (p. 228) creating a greater problem of identity formation with little regard for “how we know, how we learn, and how we are taught” (p. 230). The complexity of this myth of teacher as expert, then, is a “normalizing fiction that serves to protect the status quo” (p. 229).

Teachers are self-made, as the third conflicting myth suggests that teachers are “born” rather than developed. Britzman asserts that this myth, more than the others, devalues “teacher education, educational theory, and the social process of acknowledging the values and interests one brings to and constructs because of the educational encounter” (p. 230). Britzman explains that Jack subscribed to this myth of self-making, in that he described his individualistic construction saying, “I think that teaching is something that I’m going to learn how to do myself” (p. 230). Whereas, Jamie understood the pitfalls of unmediated experience; she also fully engaged herself in the practice of “teaching herself how to teach” (p. 231) thereby ascribing to the myth of the “natural teacher”. Britzman argues that in the “supposedly self-made world of the teacher, pedagogy is positioned as a product of one’s personality and therefore is replaced by teaching style” (p. 231).

Britzman states that viewing teaching style as a product is a mistaken assumption, concluding that teaching style defined as being synonymous with knowledge is a highly simplistic understanding of teacher pedagogy and practice. Britzman argues this over simplistic view reduces teaching style to a costume where one may try and try on again

until one finds the right fit. “Such a metaphor reduces pedagogy to its most mechanical moment” (p. 231). Britzman maintains the complex nature of teacher pedagogy and practice is filtered through teacher identity formation and the social, cultural relations “among teachers, students, school culture, and the larger social world” (p. 232) and that for student teachers, like Jack and Jamie, the need for unfettered educational discourse is essential in the development of their teacher identities.

Britzman’s study of the lived experiences of student teachers illuminates the complex social environment of schools where teachers interact and construct their teaching selves. The findings of this study suggest that teacher identity best develops within a discursive environment where beginning teachers are afforded mentorship from their peers. "Enacted in every pedagogy are the tensions between knowing and being, thought and action, theory and practice, knowledge and experience, the technical and the existential, the objective and the subjective" (p. 2). Here, Britzman articulates specific skills which a beginning teacher must foster in order to move teacher identity apart from the myths of teacher identity formation:

1. observation, taking on the perspectives of others
2. identity of teacher as inquirer
3. teacher as researcher

This study addresses many of the important issues related to teacher identity and seeks to encourage the “disruptions of the taken-for-granted [which] may help make available discursive practices that open the pedagogic imagination” (p.240). Teachers

caught within the dominant mandates and expectations of the high-stakes testing environment find themselves within a complex tension between pedagogy and practice, where their internal discourse and professional knowledge is counter to the practices expected in the educational setting.

The findings of this study indicate that teachers' identity constructions are best fostered and explored through what Britzman terms the *complexity of relationships* where "the contradictory dynamics of their own biography ... help them determine the interventions necessary to move beyond the sway of cultural authority" (p. 232). Britzman concludes that "[e]veryone in teacher education needs the space and encouragement to raise questions that attend to the possible and acknowledge the uncertainty of our educational lives" (p. 241). Teacher identity, as Britzman suggests, is more of an invitation to the possible.

Identity and Becoming a Teacher

In a qualitative, single case study Larson and Philips (2005) discuss the influence of ideological conflict between a teacher's preservice program, the contested spaces of school experience and the resulting influence on teachers' emerging teacher identity. This study is grounded in the perspectives of Bakhtin (1981), Foucault (1972), Weedon (1987), Britzman (2003), and Lather (1991). Choosing to use poststructural feminism as the primary theoretical framework for this study, Larson and Philips position their research "in the midst ... of student teacher's becoming" (p. 312) as she struggles with the colliding and conflicting discourses of a highly scripted district literacy program and the comprehensive literacy (Goodman, 1996; Halliday, 1975; Smith, 1994) discourse

taught in her university courses. Overall, this study seeks to disrupt the “silent regulation” (Luke & Gore, 1992, p. 4) of developing teacher identities through mandated, scientifically proven and scripted programs.

Larson and Philips begin their research with a group of preservice teachers who were to teach two days a week in an after-school program for at-risk students during their full-time teaching experience. The task for these preservice teachers was to create literacy experiences which were rich, differentiated and collaboratively planned for a highly impoverished socio-economic school setting. These preservice teachers received training before hand in the scripted reading program. Realizing early on that fidelity to the script was inherent to the “success” of this type of program, the researchers explain that they intervened on behalf of the preservice teachers. Larson and Philips expressed concerns of these preservice teachers acting as stringent “script caretakers” of a scientifically proven program rather than engaging as teachers of a rich, comprehensive literacy experience.

However, negotiations on these issues with the after-school program administrators failed and the preservice teachers were withdrawn from the program. Interestingly, the researchers saw this “failed” project as a bend in the road where, although they had decided to abandon the initial research intent, they chose to follow the data as it unfolded. While Larson and Philips began this study to examine the practices of a group of preservice teachers who volunteered to teach in an elementary after-school literacy program serving “at-risk” children, they specifically turn the research focus to one student teacher, Claire. Larson and Philips state:

we want to deconstruct how the authoritative discourse of the scripted reading program and our own discourse of comprehensive literacy struggle at the site of one student teachers subjectivity to form her emerging understanding of reading and reading instruction. (2005, p. 317).

As a result, Larson and Philips outline three revised purposes of this inquiry:

1. to illustrate how colliding discourses conflict with pre-service teachers' emerging identity
2. to highlight and consider the possibility of agency and hope, and
3. to analyze the dynamics of the researchers own positions and roles.

In sum, this “inquiry becomes a study of subjectivity and of authoritative discourses shaping ourselves and our students; it is work situated in a contested space” (p. 311) where the researchers “are not neutral voices, but active respondents in this study (p. 313). Stating their overarching ideological beliefs, Larson and Philips articulate that while they are conflicted in the telling of this study of one pre-service teachers' experience they are hopeful and propelled with “good intentions” (p. 313) among these discourses of power, positioning and agency.

Data collection occurred over a five month period with an initial volunteer study sample of 6 (5 females and 1 male) graduate teacher education students attending a private university. Respondents had previously taken two literacy courses taught by the researchers of the study, which were based on comprehensive literacy discourse practices. From the six initial respondents, three students identified themselves as being available to continue to participate in the study. Preliminary data were collected on all three teachers' respondent experience; a single case study approach was chosen to

illuminate data from one single preservice teacher's experience. Data sets were collected utilizing four audiotaped and transcribed meetings (45 – 90 minutes), one site observation, email communications, and teacher educator reflective journals.

Two primary discourses were identified to categorize and code the data: that of a mandated reading federal policy and of the researchers reading course theoretical framework. Data were read and coded based on “how this language reappeared and was used by the student teachers” (p. 314) guided by Britzman's (2003) questions to express and articulate the struggle of teachers' voices within conflicting discourses. One preservice teacher's language emerged in the data as being more reflective and representative as a leader as she questioned and made connections in her struggle between the two discourses. This single case was chosen to highlight.

The preservice teacher highlighted in this study presents as being highly conflicted between the authoritative discourse and the discourse of her university class experiences. “This in turn played upon her subjectivities and her emerging identity as a teacher of literacy” (p. 317). Each time Claire posted to her emails she was pulled and pushed by her thinking and demonstrated “dramatic shifts” (p. 317) as she mediated these discourses of power. Struggling between the authoritative discourses, Claire expresses her conflicted experiences while she worked to negotiate the discourses of power. Claire recognized that she was positioned by blame and pulled into participating and helping with the after-school program when she dropped by to see how things were going. The neutralizing effect that the authoritative discourse had on Claire's teacher identity was profound. Claire

felt she “was not qualified to disagree with the scientific evidence” of the program. Later Claire described herself and her colleagues as “nobodies”, without ideas, experience, or expertise” (p. 318). Claire lacked all ability to see the agentic moments as she was consumed by the complex, ideological force of the mandated program.

Based on the findings of this study, four unresolved elements were fundamental in influencing teacher identity and understanding the pedagogical and practice decisions made by this teacher. First, the stark differences of the two programs created an either/or binary, forcing teachers to choose a right or wrong. Secondly, it is important that teachers engage in discursive negotiations of pedagogical and practice choices. Third, this teacher recognized the need to deliberately utilize the skills of critical literacy and deconstruction. Fourth, the “power of student collaboration” (p. 322) came to the forefront in relation to identity where we might “continue to experiment with ... group spaces of inquiry and mentorship that offer alternative discourses to frame student teaching experiences” (p. 322). The narrative nature of this study is a compelling “story” of teacher identity in relation to conflicting pedagogy and practice. Larson and Philips conclude, “[A]s privileged discourses sanctioned by government move to mandate the identity of pre-service teachers, perhaps such inquiries become even more urgent” (p. 322) within the current high-stakes testing environment.

Identity and Courage

Telling the “miseducative” (Dewey, 1938) story of one teacher (Naomi), “a story in which impossible contradictions, gaps, and silences are named,” Huber and Whelan (1999) explore this narrative through the conceptual framework of the lived “professional

knowledge landscape” (Clandinin & Connelly, 1995). Following a qualitative methodology, this 18-month long study takes on a storytelling approach as one teacher’s story is told as “story to live by.”

A group of five teacher co-researchers acted as a “conversational place” for sharing and unpacking Naomi’s story, “situating this inquiry within a narrative conceptualization of teacher identity and the professional contexts in which teachers live and work”(p. 382). Naomi and this group of co-researchers met together to “each share stories of narrative understanding of teacher knowledge and identity.” This narrative process sought to join this group together, positioning all within a relational understanding of the contexts of each member of the group.

Within the procedures of narrative inquiry, the group heard Naomi’s story. Naomi is from a junior/senior high school in a western Canadian province where she tells of her struggles experienced while negotiating the “issues of integrating students with special needs into ‘regular’ classrooms”(Huber & Whelan, 1999, p. 382). She describes her feelings of marginalization, as the only teacher of a specialized program Naomi expressed feelings of loneliness and isolation. As Naomi continued to tell her story she began to engage in deeper, reflective processes describing the changes she experienced over time where eventually she “became an outsider” (p. 383) within the school landscape. She explains that by not following the “status quo” of the school landscape she was slowly marginalized until she found herself on the outside. Naomi unpacks this realization by saying:

I think I initially started to go there (outside of the school story), maybe not consciously, but I think soon it was a conscious decision and I was not prepared to be there in any other way ... , I think it was the only way that I could make sense. It was the only way that I could exist (p. 384).

Within Naomi's story a secondary character, Brian, is introduced as a positional power holder within Naomi's struggles. Brian is another special education teacher, who in his position of power, was responsible for the placement of each special education student. Naomi believed that Brian lived distantly and within a story which was both complicit and fully compliant with the school landscape. The principal of the school is introduced as a character who sought compliance and was disinterested in acknowledging or working through tensions. As a character of power he dismissed Naomi's concerns of her students, "eventually telling her she must either support Brian or say nothing at all" (p. 385) of her feelings and position of difference.

Another story character is introduced to the narrative, Laura, a special program needs aide, who worked alongside of Naomi. All three characters (Naomi, Brian and Laura) were present at a professional meeting where, in the later telling of this exchange, Naomi came to realize that she and Brian each storied their participation and interactions very differently. Naomi's story is filled with struggles and conflicting expectations of compliance concerning her role and obligations within the school landscape. She describes how she became increasingly aware of the "aura of silence" (p. 384) as she was slowly moved to the outside and excluded from meaningful professional conversations. Naomi's story closed with a deep sense of mourning, her "deeply felt sense of

marginalization, shaped by the conflicting nature of the stories being lived and told on her school landscape, ultimately led her to leave her community ... and position with the district” (p. 385).

Data analysis was explored drawing from Clandinin and Connelly’s (1995) conceptual framework of the “professional knowledge landscape” which views storied lives as shaped, and being shaped within the dynamics of the landscape. A group of five co-researchers acted as a backdrop to the narrative on which the story was told, retold, and told again. Huber and Whelan describe this exploration of the story as being “profoundly educative in that through the sharing of this story, the meaning was reshaped from beginning images of hopelessness to those of possibility” (p. 382). Findings from this study suggest that teachers shape and are shaped by the professional knowledge landscape in their schools. Further to this perspective, teachers caught in highly prescriptive educational settings find themselves negotiating the theory-practice tension which Clandinin and Connelly describe as a “split existence”. Teachers, like Naomi, are faced with the dilemmas of power, positioning and agency as they negotiate self and struggle with those mandates that are “being scripted for us on school landscapes” (p. 381). Importantly, understanding teacher identity as story to live by, calls for a relational understanding of teacher identity and the contexts in which teachers work.

Identity and a High-Stakes Test

Exploring the implications of teacher identity development of one teacher who sought to navigate and mediate her work as she struggled within the high-stakes testing

environment, Assaf (2008) provides a deeply compelling case study. Utilizing ethnographic and grounded theory methodologies Assaf asked:

How did the professional identity of one reading specialist shift in response to the testing pressures at her urban elementary school?

This study examines the “complexities and contextual tensions” (p. 239) one reading specialist encounters as she navigates her own teaching identity within a teaching environment responding to the mandates of high stakes accountability. Framing this case study, Assaf draws upon identity theory and the seminal research on high-stakes testing. Sociocultural theories of identity frame Assaf’s study within a “socially constructed, complicated, fragmented, contradictory and fluid” (p. 240) perspective. The literature framing this study focuses on the pressures of mandated high-stakes testing on teacher enacted pedagogy and practice, and the situated influences and outcomes of the high-stakes testing environment experienced by one urban elementary school teacher.

Examining one teacher’s (Marsha) struggle within the “competing and conflicting forces” (p. 240) of high-stakes testing pressures, Assaf illuminates how these high-stakes testing pressures powerfully thwarted and impacted her professional knowledge. The teacher, Marsha, was purposely chosen because she was considered to be a highly qualified teacher and experienced reading specialist within her educational setting. Marsha had a Master’s degree in literacy instruction, certified as a reading specialist and English as a Second Language Teacher and had worked in the education field for over 37 years.

Assaf recognizes that Marsha interacts within a highly agentic pedagogic and practice setting where she must make decisions between her professional knowledge beliefs and the pressing mandates of the high-stakes testing environment (Texas Assessment of Basic Skills). Assaf describes a highly complex view into one teacher's struggle to support her students and their academic success measured by the test and her own experienced teacher identity. Marsha's poignant description of how the drop in her school's test scores dramatically affected her deeply rooted professional knowledge and beliefs and "how her professional identity shifted in order to accommodate testing demands" (p. 240) ultimately moving her to align her pedagogy and practices with the test items on the test.

Assaf articulates Marsha's shift in pedagogy and practice as being directly related to the instructional decisions and practices Marsha felt compelled to make to ensure her students would achieve acceptable test scores. As test scores waxed or waned Marsha responded directly to the test and aligned her reading instruction to those curricular items being tested. Marsha's teacher identity shifted alongside of these tensions and decisions she was forced to make in her literacy practices. Outwardly, Marsha described a sense of competing tensions as she "grappled with how to stay true to her own professional identity" (p. 239) and work within the situated complexities of her work environment.

Findings from this study suggest that high-stakes testing pressures at Marsha's school affected both her pedagogical and practice and impacted her professional identity. Assaf concludes that when high-stakes testing "can influence teachers" responsibility and

ethical sense of what they should do for their students and who they need to be as teachers” (Assaf, 2008 p.239) the consequences of high-stakes testing pressures have moved far “beyond basic test-focused practices” (p. 250).

Identity in Review

From a sociocultural perspective, the literature reviewed suggests that teacher *identity* is a highly local, situated, and fluid construct. From a historical perspective, teacher identity has continued to be recognized and brought to the forefront of the conversation regarding seminal influences on teacher’s work. From a pedagogy and practice perspective, additional studies examining effective teachers, teacher quality, teacher preparation and teacher professional development are essential to developing an awareness of the complexity of teacher identity within high-stakes reform settings. The findings from these studies reviewed suggest that identity, while highly relational, shapes and is shaped by social interactions, and power relations inscribed within the educational setting (Street, 1984).

Viewed in this way, identity is enacted as we draw from the *self* in spaces, times, and relationships (Moje, 2004) and from the certain ways that our identity is recognized or not (Gee, 2001). Flores and Day (2006) characterize the complexities of teacher identity construction as “the interplay between contextual, cultural, and biographical factors” (p. 219). Teachers working within the high-stakes testing environment work within competing and conflicting tensions where power, position and agency become the central prospects to who teachers need to be. The common theme among these studies

reviewed is summed by Britzman's view that "there can be no learning without conflict" (Britzman, 2003, p. 3). This struggle acts as a catalyst for change and development and involves a process of negotiation in which mediating influences and socializing agents influence identity as it is enacted.

Identity, therefore, existing within a political, social, historical, cultural, and ideological context, is a multi-faceted construct that values specific ways of constructing identity, while silencing others. Power viewed within a sociocultural lens is an embedded element of identity forming, where identity as a fluid, shifting construct is recognized differently by others. Identity is unstable as it exists in relation and is situated and positioned socially and culturally. Johnston (2007) contends that this is the agentic narrative, a moment of invitation to generate change. Power and identity from this position undergirds the "fields" (Bourdieu & Passeron, 1996) of identity development and construction, offering a central mediating place of agency where teacher identity is both unstable and generative. From this perspective, the notion of multiple teacher identities which are local and situated connects strongly to the pedagogical and practice choices of teachers.

Here, Dewey's (1938) notion of "situation" and "experience" describes the tenuous place of high-stakes testing within a temporal climate where teachers work to navigate and articulate their teacher *identity*. Central to Dewey's work is the recognition for a deep, shared understanding in articulating the complex characteristics in teacher identity formation. The need to consider the heteroglossial (Bakhtin, 1981) nature of

relationship between personal and professional identity, and how identity shape, and are shaped by the wider discourse within the community of practice (Wenger, 1998) becomes much more than ideological difference. This space is filled with complexity and contention and involves reflecting upon both the internal and external factors, which shape our identity as teachers.

Britzman (2003) believes teacher identity is formed within the tensions and complexity of those pedagogical relationships. How we see, and how we are seen, “form the narratives, metaphors, and philosophy statements” (Alsup, 2006, p. 7) which root our very basic epistemological beliefs. The relationship here to pedagogy is rooted in our identity and enacted in our practice. These issues are negotiated, and renegotiated as teachers strive to express their identity within a confluence of influence.

The complex character of power, positioning and agency within an enacted teacher identity is central to developing a critical understanding on one’s identity within the spatial relations of high-stakes testing environments. Understanding one’s enacted teacher identity begins with a deep awareness and recognition of the effects of high-stakes testing on teacher pedagogy and practice. Teachers must negotiate and mediate often conflicting views regarding pedagogy and practice in order to make sense of the teacher identity they choose to enact (Lortie, 1975). Education researchers will be well served to continue to explore and study the relations attributed to identity and its influence on teacher pedagogy and practice in the high-stakes testing educational environment.

High-Stakes Testing through Diverse Lenses

The purpose of the second section of the review of literature is to discuss seminal educational research focusing on the effects of high-stakes testing on teacher pedagogy, practice and identity. The body of work reviewed encompasses a diverse selection of research methodology, exploring effects attributed to high-stakes testing. This section of the review will first examine proponent and opponent views of educational research. Secondly, this section will present a diverse representation of methodological diversity utilized to study the effects of high-stakes testing.

The initial search of the literature found that within the field there exists an abundance of literature on the topic of high-stakes testing, although few hold to research standards of quality (Howe & Eisenhart, 1990). Using standard search terms, such as high-stakes testing, mandated tests, accountability reform, teacher's perceptions, etc., produced a large amount of scholarly writing on this subject. Further focusing the search of the literature for this review, studies which utilized a diversity of methodologies and which maintained standards of research quality were searched; this parameter reduced the works to a more selective body of both qualitative and quantitative research. Methodological and analytical diversity was considered in relation to the central research question. Again, each research study was considered and selected through the lens of Howe and Eisenhart's (1990) standards of quality and the strength of the study to support the central research inquiry.

High-Stakes Testing: Proponent and Opponent Views

Proponents of high stakes testing argue that these tests convey student achievement and act as an effective lever and monitor of system effectiveness. Early advocates of the high-stakes testing movement promised that high-stakes testing would move America's schools toward achieving high academic standards and accomplish what schools had been unable to in the past – educate all students, regardless of social, economic or racial status (Paige, 2001; Ravitch, 1995). In response, Smith and O'Day (1990) proposed that standards-based reform would set in motion a systematic reform movement which would include professional development and other conditions to develop teacher professionalism within a testing environment. Among the supporters of high-stakes accountability, Hess (2002) contended that for high-stakes testing to have a significant effect on overall educational outcomes and teacher quality, “educators must be rewarded or sanctioned on the basis of student performance” (p. 73). Citing the highly controversial results of the early Texas Assessment of Academic Skills research and supported by an assumed public commitment to standards-based reform, Hess argued for rewards and sanctions to be implemented as effective levers for incentive and change.

Additionally, in a recent Public Agenda survey conducted by Farkas, Johnson and Duffet (2003), teachers reported that they supported standardized tests for student promotion and that, in their opinion, these standards of achievement improved student academic performance. Further to this perspective, Roderick, Jacob and Bryk (2002), analyzing the impact of high-stakes testing, reported higher levels of student achievement attributed to the implementation of high-stakes tests in Chicago schools. Citing their

findings from a 50-state analysis, Carnoy and Loeb (2002) provided a perspective view which described positive relationships between mandated accountability measures and overall performance of student achievement on high-stakes tests.

In support of this view, educational researchers acting as proponents of high-stakes testing argue that high-stakes tests are a highly efficient and necessary vehicle to hold schools, programs, and larger state initiatives accountable, reward those that are high performing, and identify those failing so they may be targeted for extra support (Firestone, Monfils, & Camilli, 2001). In a study by Stecher, Hamilton and Gonzales (2003) the positive effects of high-stakes tests in relation to students being better informed about their knowledge and skill levels, motivating students to work harder, setting specific study skills and aligning student effort and motivation to rewards were reported. These researchers concur that accountability measures in the form of high-stakes tests yield increases in individual motivation and academic achievement while offering stringent accountability measures for assessment of students, teachers, programs, schools, districts, and federal initiatives.

Together, these proponent views suggest high-stakes tests act as a reliable and valid barometer of how well schools are doing at implementing policy and focusing the practices of teachers and school administrators by directing their work to a curriculum which is directly tied to high-stakes test content (Firestone, Goertz, & Natriello, 1997). Regardless, for those researchers who argue that high-stakes testing is both an efficient

and beneficial way to test, reform, and track educational settings a greater understanding of the complex effects of high-stakes testing is required.

In contrast to the proponent view, opponents of high-stakes testing see mandated tests as a dangerous move to control, narrow and inhibit student achievement (Corbett & Wilson, 1991). Hillocks (2002) discussed how mandated high-stakes tests adversely impact instruction and underscore many of the issues and concerns of utilizing high-stakes tests to drive curricula. Mathison and Freeman (2006), stating findings from a recent study, reported that teachers felt they had lost control over their pedagogy and practice, further stating that high-stakes tests had dramatically altered their priorities and purpose in teaching. Numerous studies have reported the narrowing of curriculum to test-based content and basic skills (Haney, 2000; Lazear, 2006; Smith, 1991). For example, in a recent study of a large scale curriculum project, Schultz and Oyster (2006) found that teachers expressed they had little or no control over the curriculum and instructional choices that mandated their work. Further to these findings, researchers Hong and Youngs (2008) and Shepard (2000) expressed concerns that the current press towards greater accountability and high-stakes testing leads to narrow, test based pedagogical and practice decisions which are highly attributed to standardized measures.

Examining the relationship between high-stakes testing and teachers' beliefs and practice, Cimbricz (2002) reviewed the current professional literature and concluded that high-stakes tests had an overall negative influence on teachers' daily pedagogy and practice. In a recent review of educational research conducted by Solorzano (2008) issues

of pressure on teachers to raise test scores, fear of state take over, and loss of curricular control within a highly test driven accountability system were found to be the norm. According to qualitative research presented from an interview study of urban educators Crocco and Costigan (2007) concluded that high-stakes testing had such a “negative effect on beginning teachers’ perceptions about their opportunities for developing a satisfying teaching practice” (p. 514) some teachers were already considering leaving the teaching profession. These teachers expressed making pedagogical and practice choices to “tow the line, for fear of retribution” (p. 529) as they mediated the demands and test pressures of the high-stakes testing environment. In contrast, stating findings from a large scale experimental study conducted in Ireland during the 1970’s, Kellaghan, Madaus and Airasian (1982) found that teachers working with standardized tests which had no consequences attached reported no anxiety related to their teaching. These researchers stated that the findings confirmed their observations that the source of anxiety does is not directly attributed to the test but rather with the consequences.

Drawing upon a longitudinal field-based study, McNeil (2000) claimed that standardization, central controls and high-stakes testing “harms teaching and learning” (p.xxvii) where negative shifts in the cultural practices of teachers in Houston’s public magnet schools resulted in a narrowing of the curriculum to the test. In a more recent mixed methods study McNeil, Coppola, Radigan and Heilig (2008), utilizing both archival data sources and ethnographic case study methodology, report how high-stakes testing practices “individually, collectively, and cumulatively work against” (p.7)

effective teaching practices. McNeil and her colleagues traced the “pressure to comply with a highly standardized accountability system over several years ... show[ed] that even a well-meaning faculty eager to improve the school can be rendered ineffectual by the short-term pressures to produce numbers” (p. 25).

Additionally, findings from a large scale survey indicate that teachers were compelled to direct their teaching pedagogy and practice to test preparation resulting in a diminished sense of teacher autonomy and professionalism (Abrams et al., 2003; Sleeter & Stillman, 2007). While these researchers argue that high-stakes testing policies and the standards movement are designed from a flawed assumption of human motivation (Kellaghan, Madaus, & Raczek 1996; Urdan & Paris, 1994), together they highlight a central paradox of the high-stakes testing environment. Here, high-stakes testing describes a “carrot and stick” metaphor (Allington, 2002) where the stakes attributed to tests are used for purposes far beyond their intended construct. Regardless of perspective or interpretation, both proponent and opponent groups provide convincing data to support their position in relation to high-stakes testing. However, within the field educational researchers find other’s research stance on high-stakes testing to be antithetical to their own reasoning.

Research on the effects of high-stakes testing reveals that testing can have intended or unintended effects; consequences of these tests play out as both subtle yet powerful influences within the educational setting. Some educational researchers argue that the effects of high-stakes tests are a useful and highly efficient means of determining

educational practice, obtaining achievement indicators and motivating educational change (Carnoy, 2005; Carnoy & Loeb, 2002; Coyne & Harn, 2006; Green & Winters, 2004; Swanson, 2006). Other researchers advocate that high-stakes testing is directly attributed to negative educational effects of curriculum narrowing, teaching to the test, promotion and retention policies, test score corruption, and school rewards and sanctions (Allington, 2002; Hillocks, 2002; Hoffman et al., 2001; Johnston, 1998; Linn, 2003; McGill-Franzen & Allington, 2006; Nichols & Berliner, 2007; Paris & McEvoy, 2002; Sirotnik, 2004; Smith, 1991). These perspective views of proponent and opponent research groups maintain that the use of standardized testing for federal, state, or local accountability both influences and impacts educational practice.

High-Stakes Testing: Methodological Diversity

This section of the review of literature focuses on how researchers have examined the issue on the effects of high-stakes utilizing diverse research methodology. Drawing from a large body of research examining the issue, effects of high-stakes testing on educational practice, four studies were selected to present a diverse representation of theory, research design, methodology, and focus. Inclusion in this review was determined on issues of research quality, relevance, and diversity, resulting in a representative review of the growing corpus of studies addressing the effects of high-stakes testing on education. Moreover, these studies reflect the intent of the central research questions and describe the effects of high-stakes testing and illuminate how those effects transfer to teacher pedagogy, practice, and identity as teachers work within the high-stakes testing environment.

Curricular Control: Metasynthesis

Au's (2007) qualitative metasynthesis examines current research related to high-stakes testing and curricular control. This study analyzed 49 qualitative studies asking:

What, if any, is the effect of high-stakes testing on curriculum?

Au (2007) contends "[a] test is high-stakes when its results are used to make important decisions that affect students, teachers, administrators, communities, schools and districts" (p. 258). Utilizing template analysis, Au identified two terms to focus his analysis, "curriculum" and "high-stakes." Au followed this procedure through a two-stage development to a final template for interpreting the textual data. Choosing to use a qualitative methodology of metasynthesis permitted Au to synthesize a large body of research studies in order to gain a broad and more general view of the phenomenon (Thorne, Jensen, Kearney, Nobli, & Sandelowski, 2004). Here, it may be argued that this summing up of a large body of evidence may be somewhat antithetical to *synthesis* and thus, may stand as an internal threat to the validity of Au's study. Drawing on data from individual studies, which are highly temporal in nature and influenced by the ecology of the local context of each study in order to determine a "snapshot and general description", (Au, 2007, p. 262), is problematic. Researchers have stated that synthesizing others' research may over generalize discreet and contextual findings, while drawing different research methodologies together under a generic, singular view of qualitative research (Scruggs, Mastropieri, & McDuffie, 2007). In recognition of this concern, Au explains that "a way in which the findings of this qualitative metasynthesis

are limited relates to the time periods reported on” (Au, 2007 p. 262); each report frames a different time period, influenced by contextual variables and thus may be highly contentious to over generalize findings within a metasynthesis. Au’s discussion of the temporal characteristic of metasynthesis as a possible limit to the study, stands as a refining quality and contributes to the power and generalizability of the findings and is viewed as a strength of the study.

The findings from this metasynthesis determined several contradictory trends among the 49 qualitative studies analyzed. This study revealed in a “small number of cases, high-stakes testing was associated with an increase in student-centered instruction, content integration, and subject matter expansion” (p. 263). These findings may further complicate the current views of the “relationship between high-stakes testing and classroom practice” (p. 258) and that further analysis of *how* curricular control contribute to educational inequality is needed. Overall, this study represents a topical review of high-stakes testing and curricular control highlighting the effects of policy on educational practice. Au concludes that high-stakes testing affects three significant characteristics of curriculum: subject matter content, formal content knowledge, and instructional or pedagogic practices. The major findings of this study indicate that high-stakes testing, does indeed, narrow curriculum to focusing on test content.

Instructional Influence: Case Study

Reporting the findings from interviews with 59 teachers and 20 parents, Barksdale-Ladd and Thomas (2000) present a call to “stakeholders in children’s

education to make known the deleterious effects of state testing to those in charge of state-mandated testing” (p. 384). Recognizing the current high-stakes testing environment resulting from the 1983 publication of *A Nation at Risk* (National Commission on Excellence, 1983), Barksdale-Ladd and Thomas state that “the nation’s teachers are now fully aware that policy and testing have essentially become one in the same” (p. 385). Barksdale-Ladd and Thomas report findings from this qualitative study asking:

What perceptions do teachers hold about mandated standards and related tests, and how do teachers make instructional decisions given these mandates (p. 386)?

Barksdale-Ladd and Thomas conducted a qualitative case study, interviewing 59 teachers, 35 teachers from a large, southern state and 24 teachers from a large, central northern state. The researchers note that the majority of the respondent sample was students in masters and doctoral-level literacy programs which “may represent a more informed sample of teachers” (p. 386).

Interviews were conducted for three teacher focus groups (two from the southern state and one from the northern state), composed of six teachers in each group; the additional 41 teachers were interviewed individually. Barksdale-Ladd and Thomas explain that the utilization of differing data collection procedures offered two perspectives: teachers in the focus group could freely express “themselves among their peers in an atmosphere in which we, as researchers, were almost nonexistent” (p. 387), while individual teacher interviews provided deeper and more specific information. The researchers suggest that the use of “both methods provided us richer data” (p. 387)

resulting in a strength of this study; conversely, this methodological choice may also be viewed as a limit.

Findings from this study determined that 75 % of teachers changed their instructional practices as a result of the impact of high-stakes testing. Describing findings that high-stakes tests have not had the effect of improving teaching the researchers highlight one teacher's response to the impact of testing stating that high-stakes tests are "giving bad teachers an excuse to continue doing what they've always done – lots of skill and drill. It's a license for bad teaching" (p. 389). Reporting teachers' views regarding the motivational push of high-stakes testing, Barksdale-Ladd and Thomas found that teachers felt increased pressure to comply and work to increase test scores. One teacher responded saying, "[t]he pressure is on. I feel pressure partly from the constant memos, I internalize the pressure, and it is always with me" (p. 390). Teachers identified their experience of having less autonomy and control over their educational choices, further stating that with the highly prescribed testing environment "[t]hey don't need real teachers to prepare children for tests and, in fact, I think they could just develop computer programs to do this" (p. 392). Barksdale-Ladd and Thomas conclude that although "teachers agreed for the most part with the need for higher standards, dissension was evident about the value of standards" (p. 389), the resulting test preparation for state tests, concerns for their students, and overall anxiety related to job security.

This study concludes with a compelling discussion of policy and implications stemming from teachers' and parents' views. Barksdale-Ladd and Thomas draw together

findings from teachers and parents across two states informing the question of what is at stake in high-stakes testing - resulting in a call to action “to make known the deleterious effects of state testing” (p. 384).

Student Proficiency and Outcomes: Time Analysis

Researchers interested in the effects of high-stakes testing on student learning will find the Amrein and Berliner (2002a) study compelling. This widely referred to comprehensive study asks the question:

Do high-stakes testing programs promote the transfer of learning that they are intended for? (p. 18)

Amrein and Berliner conducted a “time analysis” study that used archival, comparative data (National Assessment of Educational Progress) from 18 states with high-stakes testing policies, which utilized graduation exams for promotional purposes. Amrein and Berliner state “the intended outcome of high-stakes testing policies promoted throughout the nation” (p. 1) is to promote the transfer of student learning. The assumption of this study was: if mandated state tests were of benefit to learning, naturally the effect would be increased scores on the state high-stakes graduation tests.

Data collection was drawn from four different standardized tests: SAT, ACT, AP Exams, and the NAEP. Amrein and Berliner’s choice to use these measures as indicators of student’s learning has been viewed contentiously as a methodological weakness of this study (Braun, 2004) for several reasons:

1. Students who take SAT and ACT tests are not wholly representative of the national school demographic.

2. ACT and SAT measure college-bound students.
3. AP Exams are directed towards accelerated learners.
4. NAEP is not a test that students generally prepare for.
5. NAEP uses a multi-stage random sampling technique for each state.
6. Schools often “select” students to be excluded from the NAEP.

Using a time analysis technique for the statistical analysis, Amrien and Berliner found that there were little or no gains overall on test scores. In fact, this study found that “poorly conceived state tests are actually responsible for a dumbing-down of the curriculum, leading to poor performance” (Conley, 2007, p. 81) on state tests. Berliner and Amrein state that the Heisenberg Uncertainty Principle (Campbell, 1975) was “at work in both Texas and North Carolina” (Amrein & Berliner, 2002a, p. 37) where large distortions of data were found to exist in the reported test scores. Questioning the validity of the tests themselves, Amrein and Berliner concluded “that there is need for debate and transformation of current high-stakes testing policies” (p. 2) in light of this study’s findings. Amrein and Berliner posited that if teachers raised their curriculum standards and instructional practices, additionally motivating students to study, then test scores should increase on other assessment measures. Further to this, the rationale that trends on the state tests can be reliable and valid indicators of student learning, indicates that similar trends should present on other measures (Linn, 2000) of student learning.

This study found no compelling evidence that high-stakes tests increased student learning and achievement and “that those policies (of states having high-stakes testing policies) result in transfer to the broader domains of knowledge and skill for which high-

stakes test scores *must* be indicators” (Amrein & Berliner, 2002a, p. 54). A key assertion of this study is that high-stakes tests may in fact be adding to learning difficulties with no evidence of increased student learning. The Amrein and Berliner study stands as a marker representing much of the current debates in the research community (Raymond & Hanushek, 2003; Rosenshine, 2003) challenging the high-stakes policies and questioning the validity of the high-stakes tests themselves.

High-Stakes and Reform: Mixed Methods

Considered to be a landmark research study, Corbett and Wilson (1991) conducted an extensive mixed methods study of the intended and unintended consequences of high-stakes testing in local school testing programs in Pennsylvania and Maryland. The focus of this study was the concern for the increasing use of high-stakes testing as a policy tool. These researchers argued that while the pressure of high-stakes tests acts as a stimulus to encourage action, this consequence may be “contradictory to the intended goals of reform” (p.33).

Recognizing the metaphorical outcomes of testing reforms, Corbett and Wilson (1991) defined high-stakes testing as a kind of “reform by comparison” (p.2) where teachers, schools, districts and states were publically compared on test scores. The theory of the reform was that as comparisons were made, public pressure would increase, and this in turn would motivate teachers to a level of compliance where they would be compelled to align their instructional practices to the test. Corbett and Wilson stated:

[s]takes can become high when test results automatically trigger important consequences for students or the school system, and also when educators, students, or the public perceive that significant consequences accompany test results. Thus a formal trigger of consequences need not be built into the testing program for stakes to be high... (p. 27)

Contributing to this composite definition, Heubert (2002) later added that tests are viewed as high stakes “because they are used in making decisions about which students will be promoted or retained in grade and which will receive high school diplomas” (p. 1).

The overarching assumption of the study was that mandated statewide testing programs would instigate “change” at the local level. In response to high-stakes tests, Corbett and Wilson found that teachers directed their pedagogy and practice to emphasize content specific areas tested by the high-stakes tests. While curriculum alignment may be an action of change, Corbett and Wilson argue that this change in teachers’ pedagogy and practice may not be the intended outcome of reform. Moreover, these researchers state, “[t]he problem is that the pressure pushes schools into taking the wrong actions” (p. vii) influencing action which bears little resemblance to reform. Teachers reported that as the pressure to raise test scores increased in their educational setting, they directed more and more of their teaching to the test, “not because they believed that they were actually improving their instructional program,” but for “political reasons” (p. 104).

Findings from this study indicate that respondents from Maryland, the “high-stakes” state, reported greater curricular alignment with tests, greater use and application of test scores to consequential outcomes, greater numbers of students who responded to

the serious nature of tests, greater number of teachers who reported being under stress, greater teacher preparation load, greater feelings of low efficacy from teachers, greater curriculum narrowing, greater emphasis on test scores than the process of learning, and greater dissatisfaction between curricular test focus and what teachers felt should be taught. In sum, Corbett and Wilson found that the unintended consequences of high-stakes testing were the narrowing of curriculum and changing instructional practices to align with the test in both states. Interesting here, is that in the “low stakes” state of Pennsylvania, as stakes increased at the local level, the association of high-stakes consequences increased. Corbett and Wilson concluded that the level of stakes attributed to a test was less of a characteristic of the test or policy but more readily attributed to the perceptions of the test or policy.

Taken as a whole, these researchers found that as pressures to improve test scores increased, educators reported making “changes” in local practice “not because they believed that they were actually improving their instructional program,” but for “political reasons” (p. 104). As stakes increased Corbett and Wilson contend that “change” as activity and difference vs. “change” as improvement became the status quo in both Pennsylvania and Maryland. Based on the findings from this study, Corbett and Wilson state that, although the current political view of high stakes testing may choose to view high-stakes tests a powerful lever to improve test scores, high stakes testing does little to increase student learning.

Proponent, Opponent and Diversity in Review

This second section of the review of literature examined seminal educational research focusing on the effects of high-stakes testing on teacher pedagogy, practice and identity. The body of work reviewed encompasses a diverse selection of research methodology, exploring effects attributed to high-stakes testing. First, this section of the review examined the proponent and opponent views existing in the educational research. Regardless of stance, those researchers who argue that high-stakes testing is both an efficient and beneficial way to test, reform, and track educational settings and those researchers who argue that high-stakes testing is a measurement-driven reform which has narrowed curriculum, demoralized the teaching profession and represents as a highly unstable measure of achievement, present a contradictory frame where a greater understanding of the complex effects of high-stakes testing is required.

Secondly, the diverse research methodology represented in this section of the review of literature demonstrates how researchers have examined the issue on the effects of high-stakes utilizing a variety of research methodologies. Drawing from a large body of research examining the issue, effects of high-stakes testing on educational practice, the four studies selected represent an important representation of theory, research design, methodology, and focus.

High-Stakes Testing and Quantitative Survey Methodology

The purpose of the third section of the review of literature is to present significant educational research which specifically highlights those research studies which have utilized quantitative survey methodology to study the effects of high-stakes testing on teacher pedagogy, practice and identity. As a corollary, research that examines the interactions between mandated reforms and social and cultural enactments often requires the collection and analysis of data utilizing survey methodology (Salant & Dillman, 1994). Further to this argument, Baumann and Bason (2004) remind researchers that survey methodology is “the preferred means to address a research question when it is most efficient to simply *ask* those who can inform the question” (p. 288). This research addresses the relevant group – the teachers.

Teacher beliefs and perceptions of the effects of high-stakes testing are central to the success of the current educational reforms; understanding how teachers perceive the effects of high-stakes testing in relation to their work is an integral part of the standards-based accountability reform. Understanding the effects of high-stakes testing from a broad representative sample of teachers is key to further extending much of the single case or single site research previously conducted. Hence, the third section of the review of literature focuses on research which utilizes survey methodology to study teacher pedagogy, practice and identity within the high-stakes testing environment from a variety of perspectives.

Survey Research and High-Stakes Testing

Franklin and Snow-Gerono (2007) conducted a quantitative survey focusing on the perceptions of 106 teachers working in a standardized testing environment in the Northern Rocky Mountain Region. While teachers in this study reported not being anti-testing and expressed their understanding of how testing can support both teaching and learning, 95 % of teachers reported extreme feelings of stress and anxiety related to high-stakes testing pressures. Factor analysis determined three specific variables related to pressure: administrative pressure, media pressure, and pressure from other involved parties. Findings from this survey research reports how pressure from the mandated standardized test structures have influenced and changed teachers' work. Further to these findings, Abrams (2004), reporting from a study utilizing comparative survey methodology, compared the responses of Florida teachers to teacher responses from a national survey of other states using high-stakes tests. The findings from this comparative study indicate that, although teachers across the nation report negative effects from the high-stakes testing reforms, teachers in Florida report a greater impact of effects from the high-stakes test environment. Issues related to loss of professionalism, low morale and fear of retribution are consistent throughout much of the current survey research on high-stakes testing.

Additionally, in a recent quantitative cross-sectional survey study, Hanson (2006) compared levels of teacher burn-out in teachers working in high-stakes and low-stakes subject areas, grades and school settings. Findings from this study suggest that teachers

working in high-stakes subject areas, grades and school settings are less effective in their teaching than those teachers working in low-stakes subject areas, grades and school settings. Hanson concluded that one of the most salient effects of high-stakes testing may be the impeding effect on teaching and learning of disengaged and burned out teachers. Based on research findings Hanson suggests that district administrators give serious attention to not only student scores but to the emotional toll of high-stakes testing on teachers.

Several studies have examined the relationships between high-stakes testing and teacher attitudes and morale. Jones (1999) and her colleagues employed a quantitative survey methodology to “make public the voices and beliefs of teachers” (p. 2) in regards to the effects of high-stakes testing on instructional practices, teachers’ morale and attitudes in North Carolina. As a result of high-stakes testing, teachers in North Carolina reported a significant loss of morale and a greater increase in overall pressure. North Carolina teachers reported spending almost 90 % of their instructional time preparing for the test, adding that curricular areas which are not tested have fallen off the grid. These teachers stated they “simply go through the motions, receiving little or no feedback” (p. 4) while describing fear of “such punitive measures as the loss of pay incentives or requirements to take a teacher competency test” (p. 2). Nichols and Berliner (2007) state that in two independent surveys, conducted by the Public Agenda (2001; 2002), 75 % of teachers reported that since the mandates of high-stakes testing they have experienced a higher level of job related stress. However, findings from a quantitative survey conducted

by Kauffman (2005) examining the effects of high-stakes testing on second-year teachers found that teachers new to the profession expressed feelings of “comfort by the guidance and certainty offered by prescribed curriculum” (p. 21). Kauffman concludes, that while new teachers are learning the multitude of demands on teachers’ daily work, they are less concerned with asserting their professional autonomy and find a sense of support within highly constrained curriculums. While these findings suggest that new teachers are perhaps more content being mandated by reform agendas in their role as teacher, Kauffman states that those “teachers with more training and pre-service experience may seek less curriculum guidance” than those teachers who are “surviving and learning the ropes” (p. 21).

The survey research selected for inclusion in this section of the literature review is those studies which maintain standards of rigor, yielding unbiased and generalizable findings (Howe & Eisenhart, 1990). Finally, the studies selected for this section of the review of literature represent those research inquiries which have utilized survey methodology to investigate high-stakes testing within a powerful and useful methodology for collecting data on teachers’ attitudes, experiences, behaviors, perceptions and beliefs (Nardi, 2006a). This survey research examines the effects of high-stakes testing utilizing “self-report” data of working teachers. These studies represent the lived experiences of teachers and are intended to present the current seminal research from the field.

Quantitative Survey: Corruption, Impact and Perceptions

Using survey methodology, Hoffman, Assaf, and Paris (2001) conducted a study which looked at high-stakes testing in reading. Focusing their research on the state of

Texas and the national concerns surrounding media reports indicating positive and successful effects of the standards-based reform currently used in Texas to determine consequential decisions regarding tracking, promotion, and graduation of students

Hoffman et al. asked:

How much of the “success” is an illusion that masks an intrusion of testing into good teaching? (p. 482)

Hoffman, Assaf and Paris (2001) define high-stakes tests as those which “make decisions about tracking, promotion, and graduation of students” (p. 482). Hoffman and his colleagues stated two primary concerns at the onset of this research:

(1) the hidden costs of high-stakes reform on the educational community, (2) and the negative effects on minority and low-achieving students. Texas was chosen as the research site “because the accountability system and the standards-based reform effort there have been recognized as “a model” for other states to follow” (p. 482). Outlining a brief history of the standards-based reform in Texas, Hoffman et al describe a state testing environment which has become one of the most well known accountability systems in the United States, referred to as the “Texas Miracle”. Utilizing the Texas Assessment of Academic Skills (TAAS) as the primary criterion-referenced test and claiming its miraculous success, has drawn the attention of the nation to the Texas educational system.

Initially, Hoffman et al. obtained a member list of the Texas State Reading Association (TSRA) to define the target sample; this membership of approximately 4000, comprised teachers, reading specialists, curriculum supervisors, and other leadership

roles, with many of the members holding advanced degrees. Utilizing a random selection process, survey questionnaires were sent to 500 members of TSRA. Surveys consisted of 113 items derived from two previous surveys of teachers in Michigan (Urdan & Paris, 1994) and Arizona (Nolen, Haladyna, & Haas, 1989). Surveys were sent with stamped, self-addressed return envelopes; no additional incentives were offered. Reminder correspondence was initiated after 3 weeks, to encourage nonresponders. Subsequent surveys were mailed out to members until a useable total of 200 surveys were returned, resulting in a 27 % return rate, or 5 % of the total membership. The researchers note that the resulting representative sample is a “select group of educators in Texas with both expertise and experience in the teaching of reading” (p. 484). Further, the survey respondents were mostly teachers who work with low SES (socio-economic status) and minority students, and teachers who were older and more experienced than the average of Texas teachers.

Data analysis of the 200 surveys was conducted using item-level analyses; findings were reported as individual items and as combined items. Descriptive statistics were used to report findings from composite scores while percentage and number of responses identified individual items. Qualitative analysis was used to identify common themes from the final section of the survey. Hoffman and his colleagues do not further describe their analysis procedures employed in this study; this may be viewed as a weakness to the findings overall.

Findings from this study address the initial study question as well as the two primary concerns of hidden costs and negative effects on minority and low-achieving students. Hoffman et al. include statistical data as well as “teacher talk” to support their findings, further contributing to the validity, trustworthiness or generalizability of the findings. While the researchers entered this study stating overarching assumptions regarding high-stakes testing as an intrusive measure, data and analysis suggest that their initial concerns were warranted.

Stating “that the findings from this study are consistent with research on the negative effects of “high-stakes” assessments” (p. 490) affirms the study question and the primary concerns of this study. High-stakes testing undermines effective teaching leaving little time for real instruction for those minority and low-achieving students. Teachers reported that they spent 8 to 10 hours per week preparing students for the Texas Assessment of Academic Skills (TAAS). Moreover, teachers reported that they gave greater emphasis when planning their curriculum to test content and de-emphasized those areas of curriculum which were not designated as test items. These constraints on teachers’ educational practice are unintended outcomes which occurred as a result of the increased standardization, resulting in teacher choice to narrow curriculum. Hoffman, Assaf and Paris conclude that as a direct result of high-stakes testing, teachers separate their teaching practices into test content and real instruction – where in the current high-stakes testing environment there is little room available for “real teaching.”

Quantitative Survey: “Voices” from the Frontlines

Jones and Egley (2004) conducted a survey investigation into the perceptions of teachers within the high-stakes testing environment in Florida. Jones and Egley surveyed 708 teachers, who had four years experience with high-stakes testing in Florida, asking:

1. Have teacher perceptions of testing changed over the past few years?
2. Have teachers initial negative reactions against testing subsided as teachers have had a chance to work in this testing climate?

Drawing upon earlier research conducted within similar testing programs, these researchers examined the effects of testing on teachers and students. Jones and Egley surveyed a representative group of 3rd, 4th and 5th grade teachers who had four years of teaching experience within the Florida Comprehensive Assessment Test (FCAT). The sample represented 30 school districts out of the initial 67 school districts which were invited to participate in this research. Teachers were asked to complete an online survey designed to examine their “demographic information, their current teaching practices, and their beliefs about the FCAT” (p. 6). Results of the survey described a general negative perception among Florida teachers of the FCAT, with 79.9 % of the teachers reporting that the FACAT test “was *not* taking Florida’s public schools in the right direction”.

According to the researchers, teachers responding to the open-ended question discussed only the negative effects of the high-stakes Florida test. Teachers described a high level of stress related to the use and accuracy of the FCAT. Teachers expressed feeling pressure to be accountable for variables of student achievement which teachers felt were out of their control. Jones and Egley describe the high-stakes test environment

in Florida as an absolute cause-and-effect, where over use and misuse of single test scores are used to measure teacher quality and effectiveness and to hold teachers to public scrutiny.

Data from this survey study determined that the FCAT content had considerable effect on the increasingly narrow curriculum being taught. Teachers reported spending less and less time on any subject or topic outside of test “curriculum”. Teachers described an educational environment where they made daily decisions to “time and focus away from learning; and instead, placed the focus on other areas such as the tests and rewards” (p.16). As a result, teachers describe test prep in Florida as the “curriculum”. Teacher motivation has suffered since the mandates of the FCAT; findings from this study show teacher motivation as being impeded by stress, negative attitudes, low morale and loss of professionalism. These teachers reported that they believed in accountability and standards but as a result of the pressures associated with the FCAT they “reported enjoying teaching less as a result of the tests” (p. 20).

The Jones and Egley study echoes many of the current research findings where teachers “perceive their voices to be largely unheard by policymakers and complained that they had not been a part of the process of creating the accountability program” (p. 21). Accordingly, the researchers question the impetus of politicians and possible press behind many of the politically driven educational mandates, specifically the FCAT. However, while the teachers in this study reported that accountability was necessary, they maintained that the FCAT was not an effective measure to set educational goals and

measure overall accountability. Moreover, this study documents a consistent teacher voice within the high-stakes testing environment, where the “negative ... effects of testing appear to outweigh the positive” (p. 23). Importantly, Jones and Egley provide a research based forum where teachers have documented their experiences within the FCAT environment and ultimately provide policymakers with powerful insights and experiences to consider within the Florida mandated testing program.

Quantitative Survey: Mediating High-Stakes Reform

In a recent study, examining teachers’ mediation and agency within a high-stakes secondary school reform, Lasky (2005) stated a single overarching research question to guide her study:

What is the interplay among teacher identity, agency, and professional vulnerability in a context of large-scale secondary school reform (p. 901)?

This noteworthy study utilized survey and interview methodology to describe a highly sociocultural (Rogoff, 1990; Vygotsky, 1962) issue where teacher identity was in constant interaction within the context of educational reform mandates. Using a sociocultural lens, Lasky’s research is conceptually framed by theories of identity (Huberman, 1993), individual capacity (Spillane & Thompson, 1997), emotion (Dewey, 1922), vulnerability (Lasky, 2003, 2004), and agency (Wertsch, 1991).

To provide a rich data collection for this study Lasky employed a mixed-methods approach utilizing survey and interview methods. Of particular interest to this study, the survey instrument collected data on teacher background, teaching assignments, experiences with curriculum and assessment policy, opinions about high-stakes tests,

personal and professional effects of high-stakes testing, beliefs about the purposes of schooling, resource and support for reform mandates, school capacity, student engagement and learning, teacher collegiality, and school leadership. The semi-structured interviews provided a greater opportunity for in-depth responses from the study respondents about “teacher professional identity, vulnerability, and agency” (Lasky, 2005, p. 903). Findings from both the survey data and interview data reveal the simultaneous and complex nature of the ways in which teacher identity influenced and was influenced by their agency and sense of purpose within the mandated reform environment.

This study offers a Canadian perspective to the current reform movement and describes a possibly universal depiction of teacher vulnerability as teachers work to understand and experience the high-stakes testing environment. The teachers that Lasky studied expressed a deep sense of “moral purpose” (2005, p. 913) as they chose their individual routes of mediation guided by their ethical and professional values. The disjuncture that teachers expressed between their teacher identity and the expectations of the mandated reforms was one of the most significant findings from the data. Teachers told of “their unwillingness to change their identity ... (suggesting) that meditational systems may have limited influence on changing individuals long held notions ... (or) her notions of the right way to teach her subject area ... (and that) to lose vulnerability is to lose the personal relationship of teaching” (pp. 913-914). The struggle that Lasky

describes in terms of teachers being caught in the interplay of teacher identity and that of the high-stakes testing environment is one of expectations, pressures, values, and choices.

Primarily, this researcher was interested in “the ways externally generated reform mandates interact with teacher identity to affect teacher agency and their experiences” (p. 902) mediating the high-stakes reform environment. In this study, Lasky delves into significant and important research questions which explore the effects of high-stakes testing and mandated reform. Unlike many of the studies referred to in government sponsored publications, Lasky’s findings articulate the complex political, social, and economic systems which “shape school reform policy, which in turn mediates teacher identity, and teacher agency” (p. 914). Lasky’s work contributes to the systematic conceptualization required to understand the dynamic interplay of teacher pedagogy, practice and identity within the high-stakes testing environment.

Quantitative Survey: Impact and Instructional Practices

Massachusetts’s teachers responding to a quantitative survey conducted by Vogler, et al. (2002) reported that they had changed their pedagogy and practice to align with the Massachusetts Comprehensive Assessment System (MCAS) state test. This study was interested in the effects of test scores released to the public on teacher instructional practices. Data were collected from a 54-question survey instrument asking four guiding questions:

1. Have teachers changed their instructional practices since the release of high-stakes, state-mandated student performance scores?
2. In what way(s) have teachers changed their instructional practices?

3. What factors have influenced such changes?
4. Do the resulting changes in instructional practices reported by the teachers in this study correspond to current thinking about best practices in education?

Vogler acknowledges the current debate regarding the utility of high-stakes test scores in regards to being an instructional tool or motivator. Discussing the opponent and proponent views within the testing debate and drawing upon existing research, Vogler recognizes the distinction of the high-stakes tests used in the 1980s and 1990s as being markedly different than the current performance-based assessment test used in Massachusetts. Vogler describes the Massachusetts Comprehensive Assessment System (MCAS) as a state designed assessment which evaluates student learning in terms of the learning standards in the Massachusetts curriculum framework. The MCAS assessment tool assesses higher level thinking skills while previous tests have tested memorization skills (Rothman, 1995). This study examined the instructional practices of 257 teachers in light of the MCAS revised test format. Findings showed that, although the MCAS test format had changed greatly, in Massachusetts, the use of high-stakes test scores continued to act as a lever to “exert significant influence on classroom learning and instructional practices” (Vogler et al., 2002, p. 40) of these teachers.

Findings from this study suggest that teachers did indeed change their instructional practices in response to the public documentation of high-stakes test scores. Teachers expressed that the resulting changes they made in pedagogy and practice were to support the increased higher order types of test questions on the MCAS. Teachers

reported that their pedagogical and practice decisions to help students obtain higher test scores led to an overall perception of improved and better teaching practices which included those test items which addressed higher order thinking skills taken from the MCAS test formats.

Findings from the data further showed that the lowest increase in instructional changes was reported by teachers who were more experienced, having 28 years or more teaching experience. Suggesting possible explanations for these findings, Vogler postulates that while “these teachers may think that MCAS is just another fad and will soon fade away like so many other educational reforms efforts they have witnessed throughout their careers” (p. 46) they continue to position and integrate high-stakes test content within their practice. Additionally, Vogler considers that these experienced teachers may have over the years integrated these kinds of higher order instructional strategies in their practice and thus, did not report changing their pedagogical and practice choices when responding to the survey. However, the results from this study indicated that teachers reported increasing instructional practices which they deemed as being ‘best practice’ to effectively engage their students and improve their test scores.

Findings from this study present major implications for policymakers to consider. Vogler has touched on the complicated and controversial issue within the high-stakes testing environment – teaching to the test. Teachers teaching to the test may be one of the most supportive pedagogical and practice decisions a teacher can make when raising the learning of students in areas measured such as: open-response questions, creative and

critical thinking questions, problem-solving activities, writing assignments, inquiry and investigation , and problem-based assignment. Vogler’s findings lead one to consider that it is not the act of teaching to the test we need to reconsider – it appears that it is the properties of the test we are teaching to which are most worthy of our consideration.

Quantitative Survey: Negotiating What’s In and Out

In an exemplary study, Taylor, Shepard, Kinner and Rosenthal (2003) surveyed Colorado teachers asking:

What are the effects of high-stakes testing on instruction and test-related practices in Colorado?

Taylor and her colleagues were interested in determining if teachers perceived state standards or the high-stakes test to have a greater impact and influence on their pedagogy and practice. This study utilized a two-stage stratified cluster design to ensure a representative sample. Written and telephone survey were administered to 1000 Colorado teachers. Data were qualitatively coded by a survey staff and quantitatively processed using SPSS.

Findings from this study determined that “teachers voiced generally positive feelings about standards” (p.20), and in contrast voiced dissatisfaction when describing the changes they made to their instructional practices when adding curricular content and materials to align with those on the Colorado Student Assessment Program (CSAP). Teachers reported that the increased attention to the mandated state standards resulted in higher quality of teaching pedagogy and practice. Taylor, et al. state that in “almost all cases changes addressed to standards have served to make instruction more rigorous, and

teachers' detailed descriptions suggest that these changes are more than superficial" (p. 51).

In contrast to teachers' overall expression of satisfaction in terms of the standards when referring to the CSAP high-stakes test, teachers reported engaging in repetitious instruction where the goals were not about learning or possibility, but rather the primary message to students and teachers was that scores on the test must be improved. Teachers reported directing a greater amount of curricular time and emphasis on rote memorization to those content areas tested resulting in less time for and valuing of other curricular areas. Additionally, the instructional time given to "preparing and practicing for CSAP was not a good use of instructional time" (p. 51) and paradoxically worked to void the intent of the standards at the heart of the reform.

In terms of the effects of high-stakes tests, Taylor et al. (2003) found that Colorado teachers were making conscious choices to not teach in the high-stakes test grades. One teacher reported that although her "students always did well there was so much pressure and stress on how we would do as a school, I decided to teach art" (p. 26). While some teachers changed grades or subjects taught, Taylor et al. found that teachers were highly resistant to staying in 'failing' schools or choosing to go to a school "when they are going to be called a failure" (p.49). Based on the findings from this study, Taylor et al. question how teachers have positioned themselves as a result of the standards and testing reforms in Colorado. The pedagogical and practice shifts which teachers described went against what they knew to be as best practice and resulted in a powerful and

detrimental influence on their professional choices. Overwhelmingly, teachers were not against standards but did not support the testing mandates required for the CSAP.

Teachers reported that they felt demoralized, stressed, were required to direct considerable teaching time to the test, and had lost the element of “fun” from the entire school experience.

In conclusion, Taylor and her colleagues report that there is “both good news and bad news” (p. 51) within these findings. The good news, teachers do support and value standards. The bad news, teachers consider the effects of the CSAP to be “harmful” (p. 51) to both teaching and learning. Based on the findings of this study, Taylor et al. call for additional research to examine the motivational effects of high-stakes testing on teachers working within the current educational reforms.

Quantitative Survey: Surveying the Nation’s Teachers

In one of the largest scale national surveys in the United States, research was conducted on the issues of how teachers perceive the effects of high-stakes testing by the research team of Pedulla, J., Abrams, L., Madaus, G., Russell, M. Ramos, M. and Miao, J. (2003). This national study funded by Boston College's National Board on Educational Testing and Public Policy was conducted as a two-year study which surveyed a large nationally representative teacher sample. To determine how state testing programs are affecting teaching and learning this study surveyed teachers in 47 states. Of the randomly selected 12000 teachers who received the national survey, 4195 teachers participating in this study were regular or special education teachers from both urban and rural settings.

Teachers included in the sample were comparable to a greater theoretical population in terms of: age, race, school settings, and years of teaching experience.

Data collection utilized a survey instrument developed and modeled upon other previously conducted survey research in Arizona (Smith, Nobel, Heinecke, et al., 1997), Maryland (Koretz et al., 1996b), Michigan (Urdan & Paris, 1994), Texas (Haney, 2000), a National Science Foundation study (Madaus, West, Harmon, Lomax, & Viator, 1992) and a large-scale experiment focusing on the effects of high-stakes testing in Ireland (Kellaghan, Madaus, & Airasian, 1980).

Utilizing an 80-item survey Pedulla and his colleagues asked:

How do teachers perceive the effects of state-mandated testing programs on teaching and learning?

Teachers were asked to respond to statements about their state testing program, classroom practice, and student learning. Items focused on how state tests impacted:

1. school climate
2. pressure on teachers
3. perceived value of the state test
4. alignment of classroom practices with the state test
5. impact on the content and mode of instruction
6. test preparation and administration
7. perceived unintended consequences of high-stakes tests

Teachers were asked to document their attitudes and opinions regarding the effects of high-stakes testing policies and the resulting influence and impact on their pedagogy and practice. The Pedulla et al. (2003) study found that teachers who taught in

high-stakes educational settings described consequential effects which they perceived as being directly related to the expectations and pressures of the high-stakes testing environment.

Research has consistently reported that teachers hold positive views of standards (Clarke et al., 2003). However, 58 % of teachers responding to the Pedulla et al. survey stated that they believed their individual state standards to be effective standard frameworks, while approximately 75 % of the teachers surveyed concluded they felt the benefits of the tests did not outweigh the far-reaching costs and time required to comply with the testing programs. Teachers expressed they directed more and more curricular time towards test content and test taking skills. For example, 52 % of teachers reported that they spent less time teaching and learning in areas which were not designated as test content. Surveyed teachers stated that the mandated testing reforms actually contradicted teachers' knowledge of sound instructional practices. Overall, teachers, regardless of working in a low-stakes or high-states teaching environment, reported that they had to make significant instructional changes and modify their classroom practices to align with the mandated test.

Findings from the Pedulla et al. (2003) study identified Tennessee as a High-stakes/High-stakes state where “there are high stakes for districts, schools, and/or teachers *and* high stakes for students” (Abrams et al., 2003, p. 1). The Pedulla et al. study found characteristics such as rewards and sanctions were in policy and practice at all levels within the state of Tennessee. Teachers who worked in high-stakes states were

more likely to report their feelings of greater pressure and lower morale. While high-stakes testing may motivate and raise the morale of some teachers, the results of this study show a high level of increased stress and a marked decrease in teacher morale were frequently reported. Findings suggest that the pressure teachers experience is directly related to the grade or subject taught and the stakes attached to the test at that grade or subject.

Overall, the Pedulla et al. study examined the role of test preparation and curricular alignment in relation to teacher's work concluding that individual states must "refocus education policies to place greater emphasis on supporting and improving teaching and learning, rather than relying on a system of rewards and sanctions to spur change in the classrooms" (p. 27). Evidence from this study indicates that increased student learning as a result of the high-stakes testing policies was indeterminate. However, analysis of the data found that in school districts where there was greater curricular alignment and scripted resources there was also a high correlation to higher stakes for teachers. According to these findings, those states which had the highest stakes reported teacher perceptions of high pressures from district administration; specifically, elementary teachers reported feeling greater pressure as a result of the test than middle and high school teachers.

The Pedulla et al. (2003) study found little evidence of the inclusion of teacher voice in testing policy formation and implementation. Consequently, these researchers concluded from the findings that in the formulation of future testing programs teacher

voice holds an essential piece in assessing and evaluating if “these programs are having the intended effect” (p. 9) within educational reform movements. Pedulla and his colleagues call for future research of the effects in relation to teachers’ experiences working within the high-stakes testing environment. The research perspective of Pedulla and colleagues emphasizes the importance of understanding the effects of high-stakes testing with respect to the perspectives and lived experience of teachers working in the field.

Quantitative Survey Methodology in Review

This section of the review of the literature presented research which utilized survey methodology to examine the effects of high-stakes testing. With high-stakes testing affecting each and every education system in the country, it is essential that further research is conducted on large populous teacher samples. Utilizing survey research methodology affords researchers the ability to document the broad teacher voice and provide insight into issues and concerns that affect teaching and learning across a diverse educational setting. Survey research conducted on large populations allows researchers, policy makers and administrators to examine the effects of high-stakes testing from a broad teacher voice. The research included in this section of the review represents a seminal corpus of current and landmark research from the field.

Literature Review in Summary

High-stakes testing for accountability purposes has drawn the educational community into a reform movement where the expectation and responsibility to “change classroom practices and produce overall improvement in general education” (Pedulla et al., 2003, p. 10) has fallen to the teacher. Responding to a growing sense of dissatisfaction and unrest, educational researchers have examined the effects of high-stakes testing on teacher pedagogy, practice and identity. The existing corpus of educational research has focused on a variety of effects of high-stakes testing on instructional content (Corbett & Wilson, 1991; Jones et al., 1999), improvement of student performance (Haladyna, Nolen, & Haas, 1991; Hoffman et al., 2001), motivation and morale (Barksdale-Ladd & Thomas, 2000; Haney, 2000), test score corruption (Haney, 2000; McGill-Franzen & Allington, 2006; Nichols & Berliner, 2005), and teacher’s identity and agency within a high-stakes testing environment (Assaf, 2008; Heubert & Hauser, 1999; Mathison & Freeman, 2003; Rex & Nelson, 2004). Regardless of research stance, the research reviewed portrays high-stakes testing as a formidable force driving the current fundamental school reform (Abrams & Madaus, 2003; Madaus, 1988b; Smith, 1991). While numerous research studies have focused on the overall effects of high-stakes testing (Hoffman et al., 2001; Jones et al., 1999; Pedulla et al., 2003; Taylor et al., 2003) no research studies have utilized survey methodology to document teachers’ perceptions of high-stakes testing from Eastern Tennessee.

The research reviewed is intended to provide a diverse representation of the research on the effects of high-stakes testing; these studies suggest related findings that high-stakes testing does, indeed, affect teacher pedagogy, practice, and identity. The review of literature has presented the conceptual and methodological aspects of research in relation to the effects of high-stakes testing. Analysis of these studies found that researchers, regardless of stance, negative or positive, agree that high-stakes testing influences and effects teacher identity and educational pedagogy and practice. This review highlights the immediate and long-term effects which influence teachers' work within the high-stakes testing environment. The complexity and importance of these research studies is evident when one recognizes the multitude of intervening variables which may be manipulated or identified in a variety of methodologically and analytically powerful ways to examine the effects of high-stakes testing. Overall, the current methodological and analytical diversity represented in the research on the effects of high-stakes testing on educational practice represents a critical, faceted review of a complex issue. The reviewed research studies represent a significant contribution to the field, each contributing to the empirical literature in important and diverse ways.

Throughout the literature numerous studies have called for additional and more contextualized local research examining teacher's perspectives, beliefs, and experiences within the high-stakes testing environment (Crocco & Costigan, 2007; Grant, 2000). Recognizing that teachers are not passive respondents, Grant (2000) calls for teachers' voice to be considered in the development of any successful educational reform. In a

2002 literature review, Cibricz (2002) concluded that future research of teachers working in school settings is warranted to examine the extent to which high-stakes testing has effected teacher's pedagogy, and how the high-stakes context in which teachers work has affected their beliefs and practice. Wright and Choi (2005) finding the sparcity of teacher voices within the high-stakes testing debate troubling, stated that "[l]argely absent from this debate are the voices of classroom teachers who are responsible for implementing these policies into the classroom" (p. 4). Consequently, without an active and experienced teacher voice at the center of the high-stakes testing debate, it seems likely that the high-stakes testing mandates will continue to serve few and continue to be used as a metaphor for outcomes never intended (Sloane & Kelly, 2003). Given the centrality of the research questions of this study to the issues and concerns of the effects of high-stakes testing, it is worthy to initiate and include the research perspective of this study - where teachers' voices from lived experience are documented.

CHAPTER 3

METHODOLOGY

Chapter Introduction

The purpose of this descriptive correlational study is to examine the effects of high-stakes testing on teacher pedagogy, practice, and identity. Organized into three sections, this chapter outlines the methodology of this study and describes the systematic process of collecting, analyzing, and interpreting data. The first section provides the rationale for the utilization of survey methodology and an articulation of the research design. The second section presents the instrumentation, sampling procedures, and data collection. The third section concludes with the measurement methodologies employed in the data analysis.

Rationale

Survey as “Best Fit”

Those who understand the fine-grained aspects of research methodology may appreciate the “logistics of use” (Howe & Eisenhart, 1990) metaphor borrowed here. In choosing a methodology to answer a specific research question, researchers systematically look for the “best fit” (Strauss, 1987), in terms of warrant, purpose, application, and parameters of a scientific methodology. Howe and Eisenhardt, taking a “staunchly anti- or nonpositivist” (p. 6) epistemological view, remind researchers that one of the primary reasons that methodological “justifications are often inadequate or

unclear is due in no small measure to confusion about how best to think about standards for ... research design and analysis” (p. 2). Supporting this perspective, Nardi (2006a) contends that “[r]esearch questions must come first, and then the choice of the relevant method to study them should follow” (p.14). In choosing an appropriate research method these standards stood as an integral piece of the systematic and deliberate design of this research study. Utilizing survey design and methodology based upon Dillman’s (2007) Total Design Method (TDM), the survey instrument utilized in this study was designed to answer the following research questions:

1. What are the consequential effects of high-stakes testing on teachers’ pedagogy and practice?
2. What are the consequential effects of high-stakes testing in relation to teachers’ work and identity?

To investigate these research questions, survey methodology was utilized to identify ways in which specified variables effect teachers’ pedagogy, practice, and identity within the high-stakes testing environment. Descriptive and multivariate analyses were employed to draw out relationships and determine the relative influence of high-stakes testing in terms of the variables.

Within the justification for choosing to use a survey methodology to document teachers’ voices is a refutation of a connection to positivist or reductionist theoretical frameworks. Importantly, to this epistemological frame, Dressman and McCarthy (2004) articulate that each research methodology has epistemological strengths and weaknesses. Dressman and McCarthy argue it is essential for researchers to consider “using all kinds

of methods to understand literacy teaching and learning” (p.324). These assert that survey methodology is based on a set of unique assumptions which produces a different type of knowledge further adding to the strength and methodology of this study.

Choosing to utilize survey methodology which employs structured and semi-structured questions emphasizes the complexities of this research methodology, where the essence of both quantitative and qualitative research methodologies are both linked and rooted in difference and sameness. Social science research has debated the distinctions between quantitative and qualitative methodology, identifying strengths and weaknesses in either approach (Hammersley, 1992). Babbie (1998) asserts that,

[e]very observation is qualitative at the outset, whether it be your experience of someone’s beauty, the location of a pointer on a measuring scale, or a check mark entered in a questionnaire. None of these things is inherently numerical or quantitative, but sometimes it is useful to *convert* them to numerical form (p.36).

As Babbie describes, the bridge, back and forth, between quantitative and qualitative research is a fluid and decidedly subjective characteristic of research methodology. The conceptualization of this study falls squarely within these named paradigmatic spaces of quantitative and qualitative research methodology. Choosing to view the paradigmatic nature of quantitative and qualitative research as less of a dichotomous paradigm but rather as a most appropriate or “logistics in use” (Howe & Eisenhart, 1990) choice for the research design and purpose, Newman and Benz (1998) argue that research method choices are best viewed as parts of a continuum which describe common grounds and possible similarities.

Drawing upon characteristics of both quantitative and qualitative paradigms, Tashakkori and Teddlie (1998) utilized a pragmatic approach in which the naming of a specific method was secondary to utilizing a paradigm or worldview to guide their study and expand upon the meaningfulness of the findings. Kincheloe and Tobin (2006) call for researchers to entertain multiple perspectives within a critical bricolage of research methodologies, consciously moving from the “highly bankrupt dichotomy of qualitative and quantitative research methods” (p. 13) towards an epistemological view which is open to the possibilities of another. Further to this epistemological stance, Greene (2008) articulates the current challenge for educational research methodology is to arrive at a place where theory, world view and mental models hold equal attention. “In fact, if dichotomies are at all still useful in a modern world of concatenated complexities, it is because the tension between the antithetically conceived end points represents the important possibilities for creativity, ambiguity, paradox, uncertainty, ambivalence, imagination, synthesis, and vision” (Lipman-Blumen, 1985, p. 18). Within this research inquiry I am seeking to *understand* – positioning the data analysis within a kind of methodological *verstehen* frame, a perspective that is open to other understandings (Halfpenny, 2001).

Utilizing survey methodology to answer these specific research questions allows for descriptive and inferential analysis to be formulated from the data. An essential aspect of this research design is that the instrumentation be manageable and able to be inclusive of a sample which numerically is metaphorically representative (Stone, 2002) of a greater

teacher voice. Whereas, a qualitative research design using a multiple case study tradition may lose much detail and focus in the effort to focus on a large number of multiple cases and seemingly attempt to replicate quantitative, comparative measures (Wolcott, 1994). This study does not intend to “quantify” a “qualitative” experience but rather to empirically document and describe the experiences of a large, relevant group of practicing teachers who are best situated, in terms of knowledge and experience, to respond to the research question (Baumann & Bason, 2004; Dillman, 2007; Jaeger, 1997; Lave & Wenger, 1991) from a single researcher position.

Survey as “Logistics in Use”

Survey research offers a number of advantages to support studying relationships between identified variables and describing educational environments, while making it possible to study a range of research questions. Survey research operates within a real world setting, unlike an experimental situation, thus making generalizability from the findings to the larger educational context possible (Muijs, 2004). Generalization, in this context, depends highly upon the reliability and validity of the measure and how the research procedures are carried out and established within a “relatively unbiased and scientifically rigorous manner” (Rea & Parker, 2005, p. 7).

Primarily, choosing Internet survey methodology offers three powerful, effective and underlying reasoning to support its utilization for this study:

1. person power
2. broad educational sample
3. generalizability

The active rationale of this survey study is two-fold: to further understand teacher perceptions within the high-stakes testing environment and to give voice to the consequential effects on teacher pedagogy, practice, and identity. Central to this rationale, is a deep sociocultural understanding of the research questions which direct this study. While no single methodology is inherently better than another or will offer more quality than another, survey methodology has been chosen for this research study based on its ability to support a single researcher to collect data from a broad sample population and to answer the research questions in reliable and valid ways (Litwin, 1995).

Research Questions:

- What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?
- What are the consequential effects of the high-stakes testing environment in relation to teachers' work and identity?

These research questions are viewed as highly contemporary social issues that call for teachers' voices to be documented and heard. Moreover, grounding a survey methodology within “ a non-positivist perspective, which is to say [it] must be anchored nowhere other than in logics in use, in the judgments, purposes, and values” (Howe & Eisenhart, 1990, p. 8) that allows the respondents to contribute a broad and representative voice to the high-stakes testing agenda, ultimately being available to inform and shape future educational policy and practice.

In conclusion, this study utilizes the normed and validated Pedulla et al. (2003) survey instrument to ascertain teachers' beliefs and practices of the effects of high-stakes testing on their teacher pedagogy, practice, and identity. Finding little evidence of the inclusion of teacher voice in testing policy formation and implementation, Pedulla and his colleagues concluded that teacher voice must be a primary focus in future research. These researchers argue that to determine if "these programs are having the intended effect" the inclusion of teacher voice is vital to the formulation and success of future testing programs. Through the utilization of quantitative survey methodology, I am interested in understanding teachers' perceptions of the effects of high-stakes tests, their experiences, and how teachers enact their teacher identity in relation to the current high-stakes testing environment.

Research Design

This study makes use of quantitative survey methodology to guide a research inquiry documenting the dynamic voices of teachers currently engaged in school communities (Wenger, 1998), using quantitative data collection procedures to document an empirical and highly human perspective surrounding this study question. The research design of this study is both descriptive and correlational, emphasizing the relationships between variables (Punch, 1998). Variables, in this quantitative survey, are conceptualized as independent and dependent. Utilizing factor analysis, dependent variables were identified as:

1. school climate

2. pressure on teachers
3. perceived value of the test
4. alignment of classroom practices with the test

Primary independent variables were:

1. grade taught
2. teaching experience
3. setting
4. preparation

A 63-item instrument was developed based upon the previously validated, 81-item Pedulla et al. survey instrument (see Appendix A) used by the National Board on Educational Testing and Public Policy (Pedulla et al., 2003). For this current research, scales from the original survey instrument were utilized in their entirety. Additionally, questions outside of the scales, which did not support the intent of the research questions stated in this study, were not included in the final 63-item survey instrument. For example, questions such as those relating to gender, student grade tracking, student achievement levels, demographics of ESL and computers were not included in the final survey. In addition, a single open-ended question was formulated and included as an optional survey question of respondents. This survey research was supported by technology-enhanced survey design (SPSS 14.0) which offered a highly efficient method of collecting and managing data from a large sample population to the survey instrument.

Instrumentation

Choosing to utilize a survey methodology for this study offered a heuristic lens where links between the structure and semi-structure of the instrument provided a highly pragmatic as well as teacher narrative element to the instrumentation of this study. The very nature of this survey presents as democratic, in that, this instrument has the ability to obtain data from respondents with considerable precision and stand as a “reflection of the attitudes, preferences, and opinions of the very people from whom the society’s policymakers derive their mandate” (Rea & Parker, 2005, p. 3).

The instrumentation of this survey methodology is viewed as a strength of the study design, where allowing the respondents to remain anonymous adds to the validity and strength of the findings (Mitchell & Jolley, 2007). Anonymous survey design offers an opportunity for teachers’ voices to be documented and heard within a safe and anonymous response forum (Patton, 1990). This is further supported by Skolits (October 9, 2007), in conversation, stating that survey methods offer what a face-to-face situation cannot – anonymity. When asking research questions which are reflective of a high-stakes educational environment the consequence may be fear of retribution or sanction resulting in silenced voices. Therefore, ensuring that respondent voices are completely anonymous adds further strength to the resulting data within the anonymous survey format. In addition, utilizing an internet survey instrument further addresses other methodological concerns of possible researcher bias; here, the researcher exerts less influence and control over responses resulting in fewer ethical concerns and problems

(Mitchell & Jolley, 2007). For example, in using a face-to-face interview alternative methodology, verbal response time factor may present as an issue, where there may be a tendency for respondents to ‘fill up’ the conversational space with quick responses, taking little time to think and be reflective (Schon, 1983). Survey methodology offers a data collection procedure which, in terms of internal value, confidently honors confidentiality and privacy (Howe & Eisenhart, 1990) as well as the opportunity for individually paced and thoughtful responses.

Survey methodology affords the researcher the opportunity to access a large respondent population, which has direct experience with the central intent of the research question. Web based survey samples are often more representative than most face-to-face surveys (Gosling, Vasire, Srivasta, & John, 2004) where respondents are often sampled from a larger representative group. The importance of obtaining a broad representative teacher perspective or voice from the field is essential to describing the effects of high-stakes testing on teacher pedagogy, practice, and identity development. Specific to this research inquiry, collecting data to demonstrate the comparative effect of educational mandates and policies in relation to teachers’ perceptions of high-stakes testing, ensured that teacher opinions and attitudes were documented (Crawford & Impara, 2001) in relation to the high-stakes testing environment.

Semi-structured survey methodology permitted an inquiry method where several sub-questions were asked in relation to the central study questions and “systematically measured and scientifically assessed” (Nardi, 2006a, p. 17) further describing and

documenting the characteristics of the research questions (Baumann & Bason, 2004; Dillman, 2007; Nardi, 2006a). Echoing researchers in the field, Hakim (1987) suggests that survey methodology stands as a highly ‘democratic’ research design – where another may view the question, the data, and the findings to ensure or determine the validity of the study. From this perspective, utilizing a survey methodology for this research inquiry presents as a “logistics in use” (Howe & Eisenhart, 1990, p. 2) epistemological choice in terms of research methodology to convey the teacher experience within high-stakes educational settings.

Modeling Upon a Nationally Normed Survey

A highly regarded and nationally normed survey instrument developed by Pedulla, Abrams, Madaus, Russell, Ramos, and Miao (2003) was selected and adapted for use in this quantitative survey research. The original survey was developed specifically to study a national sample of teachers’ perceived effects of state-mandated testing programs. The Pedulla et al.(2003) survey instrument utilized in this current research reported firmly established reliability (see Table 3:01) and validity standards.

After critically analyzing the Pedulla et al. (2003) survey instrument it was determined that the original ten complete scales would be utilized for this research. Demographic and teacher profile information was designed to meet the intent of this research inquiry and was tailored to this population. In addition, an optional open-ended question was asked. This open-ended question provided an opportunity for respondents to further describe their experiences within the high-stakes testing environment, framed by

Table 3:01: Pedulla et al. study reliability

	Cronbach's Alpha
Climate	.64
Pressure	.75
Alignment	.73
Value	.79
Test-content areas	.57
Non-core content areas	.83
Activities	.91
School accountability	.89
Student accountability	.80
Teacher/Admin. Accountability	.84

the primary research question. The final survey instrument included closed-format questions, an optional open-ended question, background and demographic profile questions which were asked from the teacher respondent perspective within the high-stakes testing environment.

Construct Validity

As previously stated, this research utilized the published and validated Pedulla et al. (2003) survey. To determine construct validity, the researcher employed standards of research rigor to determine when survey questions were within the parameters of the knowledge and experience of the respondents and when survey questions were determined to be clear and not leading. A key element of this survey design and determining validity was to ensure respondent trust (Dillman, 2007) developed and conveyed through a clear message within the language construct of the survey

instrument. For this reason, demographic items were positioned at the beginning of the survey instrument used for this study. Teacher profile items were positioned at the end of the survey. This placement of survey items was a conscious effort by the researcher, to position respondents at the onset within the focus of the research question and work to encourage greater survey completion (Krosnick & Schuman, 1988). Finally, to ensure construct validity, the researcher sought the feedback of teachers to the preliminary survey instrument. After feedback and revisions, the researcher was confident that the parameters of the survey questions were within the limits of the research question and that survey questions were organized in a format which did not overextend or misrepresent the primary research question.

Survey Instrument

Utilizing a Likert response scale and closed-format items to elicit teachers' responses to the effects of high-stakes testing on their pedagogy, practice and teacher identity provided interval data which was analyzed using descriptive and inferential statistics. Limitations related to the utilization of a Likert-type summated rating scale recognizes the survey instrument requires subjective judgments by respondents, in terms of attitudes, beliefs, cooperativeness, and other important and intangible variables (McKenna, Kear, & Ellsworth, 1995). Within survey methodology, Likert-type items, which yield highly descriptive interval data, are viewed as a statistically powerful response format (Mitchell & Jolley, 2007).

Drawing upon the work of Baumann and his colleagues (2000) the researcher wrote several drafts of open-ended questions. Draft open-ended questions were given to a peer group of colleagues for critique, clarity of language and reliability of response. This group comprised two university professors, ten practicing teachers and four dissertation committee members. The final open-ended question was composed and written based upon review and response from this peer review. The open-ended question was formatted as an optional, extended response question and positioned as the final survey question.

The inclusion of a single open-ended question by the researcher was presented as an opportunity for respondents to explore a narrative telling of their lived “professional knowledge landscape” (Clandinin & Connelly, 1995) within the high-stakes testing environment. The optional open-ended response allowed the researcher to capture more detailed teacher response which included important and corroborative data to further describe teacher experience within the high-stakes testing environment (Mitchell & Jolley, 2007). Specifically, the open-ended question provided a narrative data comparative within the self-report format of survey methodology (Wilson, 2002) where juxtaposed to the closed response survey data set these respondent data presented as a highly corroborative measure. Providing both closed and open format response opportunities added to the strength of the survey instrument and the findings of the data. Moreover, viewing the teacher responses from two distinct data sources contributes to the construct validity of this survey instrument (Mitchell & Jolley, 2007) and the significance of the findings.

Survey Domains

Specific variables surveyed are inclusive within 4 domains:

1. school climate
2. pressure on teachers
3. perceived value of the test
4. alignment of classroom practices with the test

The open-ended question was surveyed as:

5. perceived effects of high-stakes testing in relation to teacher's work and identity.

The final survey instrument was piloted to a relevant group of eleven peers for appropriateness of questions, accuracy, completion time and fidelity to the primary research questions. Feedback was considered and revisions were made based on recommendations resulting from this peer review process. Piloting the survey was essential to further establishing construct validity before the administration of this survey to the defined sampling frame. Primarily, feedback from teachers completing and critiquing the test survey reported a shorter completion time and recommended placing no more than six survey response lines per online page. The final version of the Teacher Survey is included in Appendix B.

Target Population

The purpose of this research was to survey the representative perceptions of a highly demographic population of teachers who work within the current high-stakes testing environment. This research drew from a sampling frame defined by six Eastern

Tennessee county and city school systems to describe a sample that was characteristic and representative of a larger population.

The researcher was cognizant that choosing to utilize internet assisted survey instrumentation may impact the sample in several ways:

1. coverage error (Couper, 2000), limitations related to those with Internet access (Dillman, 2007)
2. sampling error (Rea & Parker, 2005), identifying the sampling frame
3. generalizability (Nardi, 2006a), inferring from the sampling frame
4. nonresponse (Groves, Dillman, Eltinge, & Little, 2002), impacted by design
5. measurement error, (Salant & Dillman, 1994) impacted by method, instrument, questioning process, and/or target respondents

Measures of quality were undertaken to recognize and address these limitations and issues to ensure the reliability and validity of this web-based survey (Nardi, 2006a).

Additionally, as Internet surveys in educational research have become increasingly common to teachers over the last decade, issues of non-response are increasingly less problematic, thereby making data collected via Internet survey methodology highly representative of the population being surveyed (Sapsford, 2007).

The context of this study is defined by the participating six county and city school systems in this quantitative survey study which are situated in the eastern region of Tennessee. School county and city systems were chosen for their proximity to the researcher's central location of Knoxville, Tennessee, as well as to maximize comparisons and to highlight regional characteristics. East Tennessee is comprised of

fourteen school systems, of which eight eastern county and city systems were initially invited to participate in this research. After introducing the research and sending information packets to eight possible participant school districts the resulting sample were six county and city systems which consented to participate in the research study.

The study sample is intended to represent a population microcosm of a larger, generalized population of teachers. Respondents within the six cluster county and city systems define a sample which is an equally representative mix of teachers practicing in rural, suburban and urban schools in Eastern Tennessee.

Following rigorous sampling procedures the researcher:

- defined the population of interest
- specified a sampling frame of possible respondents
- identified a sampling method for selecting respondents
- determined the acceptable sample size
- sampled and collected data from the target population
- reviewed sampling process

Sampling Frame

Twenty-two school systems define the area referred to as the Eastern Tennessee Region. This region is comprised of both county and city school systems. City school systems are those which operate as autonomous systems within some of the larger county districts. (For example, Alcoa City operates within Oakridge County and Maryville City operates within the area of Blount County.) These county and city system demographics

are representative of a range in socioeconomic status, rural, suburban and urban setting, cultural and race demographic, as well as leadership and professional development initiatives available. Systems identified and located within the geographic cluster and included in the initial invitation to participate in the survey research were:

- Alcoa City
- Anderson County
- Blount County
- Jefferson County
- Knox County
- Maryville City
- Roane County
- Sevier County

After obtaining regional approval, representatives were contacted from each school county and city system and invited to participate in the survey research. Within the proposed sampling frame two counties declined to participate in the research study for stated reasons such as: teachers recently mandated to participate in surveys authorized by state and county administration and concerns for close of school year administrative tasks which were overloading teachers at this time. From the initial eight county and city systems invited to participate; six county and city systems agreed to participate in the research study.

All participating school county and city systems have the properties which include the necessary characteristics to be considered as a representative sampling frame

(see Tennessee Department of Education). Representative sampling was utilized, offering equal probability of inclusion to all teachers identified within the sampling frame.

Teachers included in the sample were regular classroom teachers, special education, and specialist teachers involved with core curriculum from rural, suburban and urban settings.

Of special interest to this study were teachers within the sample who taught in grades 3 through 8; in Tennessee, these grades are considered the ‘TCAP Grades’ by teachers and administrators. Researchers examining similar frames of samples within high-stakes settings have found that teachers refer to such grades as “milepost grades” (Stecher & Barron, 2001, p. 259) where teachers teaching in these high-stakes grades report feelings of job dissatisfaction, related factors of stress and overall feelings of low morale among colleagues. For example, in a recent study Abrams, Pedulla and Madaus (2003) found that teachers working in high-stakes testing settings distinguished between curricular areas referring to reading and math as high-stakes subject areas and art, music and physical education as low-stakes curricular areas. While this group of grades 3 through 8 teachers are of special interest, the researcher has targeted the sampling design to include a representative sample of grades K-12 teachers who have direct experience with stakes attached to testing to allow for differences or correlations between groups to be validly assessed and compared (Sapsford, 2007).

Sample Size

Recognizing the importance of sample size as a rigorous component of research design, the researcher utilized rigorous sampling strategies to elicit a confident sample

size, thereby strengthening the external and internal standards of validity and the generalizability of findings. The possible sample frame of this study was approximately 2000 certified teachers of grades K-12 currently working in six county and city school systems in Eastern Tennessee.

The agreed instrumentation protocol for accessing teachers determined that the researcher must direct all research requests to the district representative who would then forward the research request on to each individual school principal. For each school the principal acted as a 'gatekeeper' in terms of access to teachers in each school. Consequently, the researcher, acting through the principal as 'gatekeeper', was not able to ensure the research request was received by the possible sampling frame of 2000 teachers. These principals, it was assumed, would then forward the research request to individual school teachers. It is noted that the researcher has no definitive data to determine whether all possible teachers within the sampling frame received any of the research requests as the decision to forward the research request to teachers was at the discretion of each school principal. While this occurrence stands as a notable characteristic of this research, all reasonable measures were employed to ensure the response rate and resulting sample was a representative population of Eastern Tennessee teachers.

Of the initial 813 survey responses, only those survey responses which were 100 percent complete were included in the sample; partially complete survey responses were not included in the data set. This parameter reduced the respondent sample size to

N= 408, representing 20.4 % of the possible target sample. The decision to include complete respondent data sets met the researcher's criteria of not including missing data for any scale or specific question, thereby further contributing to the overall reliability and validity of the findings (Fowler, 2002). Of the 408 teachers who completed the survey, 112 teachers chose to provide responses to the optional open-ended second research question, representing 27.5 % of the respondent sample.

Data Collection

Upon approval from the Dissertation Committee, the University of Tennessee Internal Review Board, county and school principal approval, a six week data collection commenced. Data were collected through an on-line survey instrument from May 15 through to June 30, 2008. Survey methodology was used to ascertain the beliefs and self-reported practices of a representative sample of teachers. Specifically, the instrument surveyed questions about various characteristics, experiences and perceptions of the respondents in the sample within the high-stakes testing environment.

The Tennessee Department of Education was initially contacted for permission to communicate with the school county and school system representatives, which comprise the central geographic area referred to as the East Tennessee Field Services Division. Administrators for each county and city system were contacted by telephone to introduce the survey research and to request permission to conduct survey research with the teachers of their designated schools. Information packages, outlining the research purpose, rationale, survey instrument, and IRB protocols, were then sent to each county

representative for their consideration and approval for the researcher to contact school principals.

Once county and city system administrators agreed to allow the researcher to contact teachers, a letter of intent was sent to all principals with a copy of the teacher invitation to participate in the research study. As stated earlier, protocols were in place which required the researcher to initiate all research correspondence to teachers through the county or city administrator. It was agreed that the district representative would in turn direct the research request on to the principals. Principals then, at their discretion, would forward the request to participate by email on to the teachers. With the administrative representative and the school principals acting as ‘gatekeepers’ to school teachers, the researcher was unable to ensure that teachers did indeed receive the invitation to participate in the proposed research study.

Each school principal received: a cover letter, the survey instrument, and detailed information regarding teacher professional resources offered for respondent incentive. Those principals agreeing to forward the online survey instrument to their teachers were asked to discuss teacher voluntary participation at an upcoming staff meeting. An important characteristic of this survey research and resulting sample is the timing and launch of the survey research. The email survey request was forwarded to schools on May 15, with all schools completing the instructional year two days following the survey launch; therefore, it is doubtful that principals had an opportunity to herald the research request to teachers in their schools. Additionally, many teachers may not have been

aware of the survey request due to the many year-end demands on their time and close of the year duties and routines. These combined factors may have impacted the final sample and data collection.

Following agreed procedures, each of the four subsequent email contacts to teachers were sent directly to the district representative, who were to forward these requests to the principals. As previously stated, the researcher had no direct email access to ensure that all teachers in the sampling frame would or did receive the survey request. For those principals who agreed to allow the researcher to initiate survey research in their schools, an introductory email was sent to teachers via the principal as ‘gatekeeper’. This email highlighted how teachers were chosen to participate in the survey, a brief explanation of the research purpose, participation incentives, and finally, that the email survey instrument would be sent to them through their school email address the following work day. Teachers were assured that survey completion would take approximately 20 – 25 minutes, with the final open-ended question formatted as an optional text format within the survey instrument. A computer link to the survey instrument was highlighted on the introductory email letter. Teachers were explicitly told they had the opportunity to end their participation at anytime during survey participation. Participation in the survey research was voluntary and it was assumed that teachers choosing to participate in this survey were acknowledging their signed and informed consent.

At the completion of the survey, incentives in the form of professional books, were offered as tangible rewards to teachers as they evoke a sense of gratitude and desire

to support the teacher professionally on the part of the researcher (Dillman, 2007). All professional books were distributed through a self inclusionary lottery and distributed via mail delivery to 55 individual study respondents. Survey respondents were guaranteed that their email address would never be used in conjunction with the data and assured that all respondent contact information was separated from the data exported for analysis.

Data collection date was set at six weeks from the initial email survey mail-out to teachers. To maximize survey response rate, the researcher utilized persistent and repeated contacts. These data collection procedures were identified in previous research as the most significant factor in improving response rates (Dillman, 2007; Scott, 1961) along with careful implementation of a respondent-friendly survey, personalizing the initial contact and subsequent correspondence, and the inclusion and opportunity for incentives (Dillman, 2007). The invitation to participate in the survey research and subsequent emails were sent out to teachers, through the principal, at predetermined intervals (Dillman, 2007):

1. May 15, 2008 (introduction and initial invitation)
2. May 20, 2008
3. June 2, 2008
4. June 14, 2008
5. June 30, 2008

Multiple contacts with survey respondents was implemented in the survey design based on the principles of social exchange theory where multiple and more personalized communications worked to ensure high quality responses and high response rates

(Dillman, 2007). A follow-up email reminder and thank you for survey completion was sent five days after the survey launch and again the following week. Additionally, one email reminder was sent every two weeks throughout the next four weeks set for data collection. A total of five email requests and reminders were sent out to district representatives and forwarded to teachers, through the principal as ‘gatekeeper’.

Accordingly, those teachers from the representative sampling frame who chose to participate in the online survey are referred to as survey respondents. Standard representative sampling procedures allowed the researcher to make inferences to a similar and larger population by sampling a group which is demographically similar to the larger population of concern (Salant & Dillman, 1998). While this survey research generated a confident sample size, $N=408$, sample size was not determined at the onset of the study. The intent of this study was not to prove or disprove a hypothesis but rather to ask the research questions to a representative sample of Tennessee teachers.

To minimize and account for error related to coverage, sampling, measurement and non-response in the discrepancy between the estimate population and the real population to be generalized, the researcher employed successful survey design and measures of rigor (Baumann & Bason, 2004; Salant & Dillman, 1994). To ensure reliability, validity and overall quality, care in the design of the research question, the survey instrument and in the overall instrumentation of the survey research has been carefully undertaken. A sampling error of 5 % was established to satisfy concerns of potential differences of the possible respondent survey participants to those included in

the study sample. Confidence levels were set at 95 % in order to strengthen the overall generalizability of findings to a greater population (Huck, 2008).

In sum, all teachers were guaranteed anonymity and confidentiality in the resulting respondent description and narrative data analysis discussed in the results of the research. Data collected from online surveys were electronically transferred to a professionally administered server at the University of Tennessee which could only be accessed through a password secured SPSS (Statistical Package for the Social Sciences) file. All respondents to the survey were anonymous to the researcher, ensured by the research protocol which guaranteed that no IP addresses or personal identifiers would be tracked or disaggregated to the individual survey respondent. Upon IRB approval from the University of Tennessee Internal Review Board, data collection continued for a six week period, from May 15 to June 30, 2008. Data is confidential and is stored on a university server which is password protected. Data will be destroyed after five years.

Data Analysis

This research utilized a survey methodology supported by a sociocultural research stance. Analysis of the data is framed by Stone's (2002) epistemological stance of numbers as metaphors. Viewing the empirical data through Stone's metaphorical lens realizes the importance of looking beyond a tally and recognizing that "[e]very number is an assertion about similarities and differences" (p. 167) of a phenomena juxtaposed to another. Stone further contends that "similarities and differences are the ultimate basis for decisions in public policy" (p. 167) and upon which many "normative leaps" (p. 167) are

made. Empirical data viewed in this way represent a metaphor for teacher's experiences, beliefs and perceptions about high-stakes testing as it relates to their work as a teacher.

Data Analysis: Survey Instrument

The first section of this survey asked:

- What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?

The data were analyzed to answer the research question using the statistical package SPSS. Frequencies (percentages), descriptives (summary statistics), and exploratory (summary statistics and displays) statistics were utilized. Additionally, relationships between pairs of variables (i.e., years of teaching experience and school climate) were explored. The characteristics studied are referred to as variables.

Guided by the primary research questions, descriptive statistics were calculated for school climate, pressure experienced, effect on classroom practice, value of tests, and impact on content and mode of instruction. Comparative statistics were calculated for grade level, setting (rural, suburban and urban), years of teaching experience and value of the test. Frequencies were computed for categorical survey items. Factor analysis, previously established in the Pedulla et al. (2003) study guided the identification of scale scores and continuous variables which supported rigorous significance and correlational testing procedures such as:

1. Descriptive Statistics – How do respondents answer back to questions in each of the sub-scales?

2. Correlational Statistics – How do the respondent sub-scales responses relate to each other and across the whole group?
3. Pearson's r Correlation Co-efficient was utilized. Alpha levels were set at .05. Additionally, a co-efficient of determination was utilized to calculate the percentage of variance accounted for in each one of the other variables. For example, this would determine how much of the variability of one factor is explained by another factor. Utilization of a co-efficient of determination technique further supported the construct validity of the measure.
4. Analysis of Variance (ANOVA) – How do different groups of respondents answer differently than those of other groups of respondents? For example, do teachers who teach grade 3 and teachers who teach grade 5 respond significantly differently to the same question?
5. Repeated Measures (MANOVA) – Do respondents answer significantly higher on some sub-scales than on others? For example, teacher responses were analyzed to determine significance, largest impact, and impact difference by specific variables.

Data Analysis: Narrative Data

The second section of the survey consisted of a single open-ended question asking:

- What are the consequential effects of high-stakes testing in relation to teachers' work and identity?

Respondent data were extracted from the survey data and analyzed using manual coding techniques (Miles & Huberman, 1994) and computer assisted technologies (Patton, 1990). SPSS Text Analysis for Survey Research package was utilized to aid in the management of the data set and to facilitate the visual representation the data. Of the 408

teachers who were included in the complete survey sample, 112 teachers provided open-ended responses to the second research question.

Data analysis followed an iterative method which resulted in a final data set comprised of five categories representing constant patterns, trends and themes present in the narrative data. These resulting categories were defined as:

1. Test preparation and curriculum narrowing
2. Teacher identity influences
3. Test use and value of test
4. Sociocultural influences and intervening variables
5. Rewards and sanctions

Respondent data were analyzed through a highly iterative methodology. Initially, the researcher read through the narrative data to develop an overall perspective of these data (Cresswell & Plano Clark, 2007); no themes, phrases or codes were assigned. Conducting a second reading of the data, the researcher highlighted phrases and reoccurring patterns across the data set. Next, utilizing a clean data set, phrases and codes were assigned to each respondent data. Keeping close to the respondent voice this iterative measure allowed the researcher to code data based upon the words and phrases of the respondent data.

To further reduce the data during the fourth phase of data analysis, the researcher reread each respondent data set to generate relational patterns and themes based on similarities and differences across the data. Next, the researcher finely examined both coded data sets simultaneously as a comparative measure to determine the most

descriptive and appropriate code to capture the voice of the respondent data. The researcher then compiled a list of all codes assigned to the respondent data sets and then worked to sort these codes into logical and conceptual groupings and categories. From this coding scheme, guided by the research question, six categories were identified. An additional examination of the data revealed that two of the categories, sociocultural influences and intervening variables, were viewed representative of a similar conceptual theme represented in the data. These two categories were collapsed to represent a single category named sociocultural and intervening variables.

During the coding process similar phrases and words were identified to construct categories which captured and represented the respondent voice conceptually in trends and patterns across the data set. This detailed coding scheme was supported by the study's overarching theoretical framework and conceptual apparatus in determining those possible groupings and greater categories as presented in the data (Miles & Huberman, 1994). Additionally, the research questions guided the coding process as the researcher highlighted descriptive phrases, identified emerging themes and categories to further reduce the data to mutually exclusive categorical data units.

Finally, data were read and re-analyzed utilizing increasingly focused and refined coding techniques as themes and categories presented across the data set. As a result of this re-analysis, the researcher was confident in the categorization of the data set. There were only three respondent data sets that did not fall into the specified coding categories. While these data were offset into an "emergent" category and not utilized in this analysis

as they did not relate to the research question and this research analysis, but are nonetheless present in the data (Wolcott, 1994). To summarize, the data final analysis determined five categories which are representative of teacher response to the second research question:

1. Test preparation and curriculum narrowing
2. Teacher identity influences
3. Test use and value of test
4. Sociocultural influences and intervening variables
5. Rewards and sanctions

Operationally, each code related semantically to the words and phrases of the respondent data (Miles & Huberman, 1994). The first category grouped teacher responses which described how high-stakes testing had affected their ability to plan and teach what they knew to be needed. The second category drew together teachers' responses which described their feelings of pressure, loss of professionalism, feelings of demoralization and their actions within the high-stakes testing environment. The third category represents teachers' views about how they value the test and how tests are used. Category four explains teachers' sentiments about the many intervening variables which impact their work in the high-stakes testing environment. The final category described teachers' references to test rewards and sanctions and how these punitive measures impact their teacher identity. These coding categories stand as "explanatory exemplars" (Miles & Huberman, 1994, p. 65) representing the complete respondent data set in response to research question two.

As a second level of analysis and data management, codes were entered into the SPSS Text Analysis for Survey Research data analysis software program. This natural language processing software, specifically designed for survey text, enabled the researcher to manage, extract value and corroborate relationships from these text responses. SPSS Text Analysis for Surveys aided in drawing together the categorized data to further re-analyze and establish the reliability checking text alignment and fidelity to the codes. SPSS was utilized to run the text data and reliably extract and categorize key concepts based on the coding scheme extracted from the open-ended survey response data. The connectedness of results of the open-ended data is depicted in a visual representation of the respondent voice.

The researcher utilized manual and computer assisted coding techniques to both manage the process and quality of the results from the open-ended survey data and to satisfy overall issues of reliability and validity. Utilizing SPSS Text Analysis software to verify the categorization of the broad themes from the data allowed the researcher to further validate the manual analysis of the data set, noting relationships among variables and building a logical and corroborative chain of evidence representative of a greater population (Miles & Huberman, 1994).

Throughout the analysis of the open-ended data, response categories and coding decisions were guided by the research questions and the language constructs of the four factors determined in the factor analysis of the scale data (Fink, 1995). This analytic process is further described by Punch (1998) where factors from the quantitative factor

analysis are utilized as theme units of data to be directly utilized, comparatively and relationally, with the narrative data. This analytic procedure stood as a further measure to enhance the overall reliability and validity of the findings. In the final phase of data analysis, categories determined were representative of respondent teacher voice portraying the range or “multiple perspectives about each category” (Creswell, 1998, p. 144).

Supported by a sociocultural theoretical lens, data were closely examined to capture both comparable responses and responses of variation. Importantly, data viewed through both an empirical and a narrative lens provide a highly metaphorical translation of the data (Creswell, 1998) where numbers and words are metaphorical representations of the respondent perspective. Miles and Huberman (1994) remind researchers that data analysis is more often “choreographed” and revised as analyses progresses somewhat akin to a responsive spiraling process. Using teacher’s words to corroborate the findings from the empirical data acted as an agent to bring teacher’s responses beyond the linguistic structures of their words to a metaphorical story-grammar (Franzosi, 2004). Interpreting the data, drawing of conclusions and recommendations was based solely on the findings from these data. Analysis and identification of themes, relationships and larger perspectives of the respondents stood to further corroborated findings from data resulting from the first section of the survey instrument.

The findings from this empirical study are reported in three ways:

1. presented in the text of the study,

2. summarized in tables, and
3. displayed by means of figures.

Utilizing a combination of these mechanisms, for statistical presentation, further fostered an understanding of the effects of high-stakes testing on teacher pedagogy, practice, and identity (Huck, 2008). SPSS statistical program was utilized to allow the researcher to summarize the narrative data in a logical and corroborative visual representation. This powerful statistical tool enabled the research to relate the findings and illuminate corroborative narrative responses to the research question.

Methodology Conclusion

Internet survey methodology provided a manageable procedural and instrumentation research format for the single researcher. This research study limited itself to the teaching population which had access to and used computer technology; thus, as with any research, generalizability of the findings to a broader population must be determined within a *proximal similarity* (Campbell, 1986) to other educational settings and interpreted with caution when looking for application and implications for policy and/or implementation. In sum, survey methodology proved to be a powerful method in illuminating the broad voice of teacher's knowledge and experience.

It is important to recognize that while survey methodology holds specific constructs which are highly aligned to the fidelity of this research question, survey methodology may also be "limited in its ability to inform us in other ways" (Dillman, 2007, p. 339). Specifically, surveys generate data which may be collected numerically

(structured survey) or narratively (semi-structured). Data may be analyzed through descriptive, inferential statistics, and qualitative analysis procedures (Baumann & Bason, 2004). Selection of the most appropriate data analysis techniques was based on the research question and the data generated from the respondent responses to the survey instrument (Dillman, 2007; Nardi, 2006a; Neuman, 2006).

To ensure the quality of data produced, two traditional statistical measures of survey quality were utilized – reliability and validity. Reliability, previously determined in the Pedulla et al. (2003) study, using Cronbach's alpha reliability test to determine how reliable a multi-item scale may be for a given population, indicated that the survey instrument asked respondents across the sample consistently throughout the measure. Strong alpha reliabilities showed that the respondents were responding consistently to the survey items. Ensuring that quality research procedures were followed at the developmental stage, as well as adhering to the stated statistical analyses, further supported the reliability of the survey instrument used in this study (Fowler, 2002).

Establishing survey validity requires evidence from several sources to determine that the survey instrument is measuring what it says it is measuring. Concurrent validity has been established by demonstrating that similar results were obtained with other validated survey instruments. The original Pedulla et al. (2003) survey was developed from other validated survey research in Arizona (Smith et al., 1997), Maryland (Koretz, Mitchell, Barron, & Keith, 1996), Michigan (Urdan & Paris, 1994), Texas (Haney, 2000), as well as a National Science Foundation study (Madaus, West, Harmon, Lomax, &

Viator, 1992). Construct validity was determined when the results obtained from this survey reflected the stated research question and related to other aspects that were expected and associated with a high-stakes testing environment. While measures that utilize self-reported survey data rely on the assumption that the respondents answer truthfully and accurately, it is almost impossible to test this assumption. Establishing validity of an instrument for a particular purpose requires a measure of understanding and reliability of the content, criterion and construct of the survey itself, as well an awareness of the limits and ways that findings and inferences may be used in useful and appropriate ways which are both valid and reliable (Sireci, 2007).

An additional consideration in assessing the quality of survey data was to look closely at whether the respondents were actually representative of the population to which results are generalized. The teachers who completed this survey are assumed to be representative to a comparable teacher population in Eastern Tennessee in terms of age, race and ethnicity, schools settings, and years of teaching experience. However, these patterns of teacher response may not be comparable to teachers working outside of the state of Tennessee due to the levels of difference associated with proficiency standards, professional development and resources. Consequently, while the results of this research are assumed to represent effects of high-stakes testing on teacher pedagogy, practice, and identity they should be interpreted with some caution and viewed within the limitations of self-reported data.

Descriptive findings from this study revealed teachers' perceptions of the effects of high-stakes testing on their pedagogy, practice and teacher identity. Paradoxically, teachers' voices are rarely heard and much less documented in the research. Teachers' experiences are often viewed as unwarranted by those who support the use of high-stakes tests. In rebuttal, Yanow (2000) argues, "To understand the consequences of a policy for the broad range of people it will affect requires 'local knowledge'- the very mundane, expert understandings of and practical reasoning about local conditions derived from lived experience" (pp. 4-5). Clearly, teachers possess this 'local knowledge' and have the 'lived experience' necessary to articulate their perceptions of the effects of high-stakes testing as it relates to their pedagogy, practice, and identity.

This research study was interested in what is essential and central to this sample of respondents within the framing theoretical underpinnings of a Sociocultural view. In conclusion, I refer back to Howe and Eisenhardt's (1990) declaration of keeping true to a non-positivist research epistemology. Quantitative methods and survey methodologies are a best fit to answer the research questions:

- What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?
- What are the consequential effects of the high-stakes testing in relation to teachers' work and identity?

Getting "rid of truth as the goal of educational research does not necessarily mean abandoning efforts to be truthful in generating new knowledge. Rather than *pursuing*

truth, however, researchers can explore different approaches to truthfulness as vehicles for reducing the ignorance of scientists and non-scientists” (Wagner, 1993, p. 22). The question, here, of justification becomes one of a prolegomenon of – logistics in use (Howe & Eisenhart, 1990).

Methodology Summary

Chapter three discussed the research design and methodology of this quantitative survey research. This study is interested in what is essential and central to this sample of respondent teachers within the framing theoretical underpinnings of a Sociocultural view. The chapter specifically presented the rationale of the survey methodology and described the research design and process of collecting, analyzing, and interpreting data. For example, chapter three has described in detail the instrumentation, sampling procedures, and data collection. This study identifies and describes seminal issues, perspectives and effects of the reform mandates of high-stakes testing through the documentation of the voices of the teachers. In sum, chapter three discussed the rationale, research design, instrumentation, sampling method, data collection procedure, and process of data analysis necessary to answer the research questions. Results of this research study are presented in Chapter four and discussed in Chapter five.

CHAPTER 4

RESULTS

Chapter Introduction

The purpose of this descriptive and correlational study was to examine the effects of high-stakes testing on teacher pedagogy, practice, and identity. Data analysis focused on the results of a survey which asked two central research questions followed by relational sub-questions:

Central Research Questions:

1. What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?
2. What are the consequential effects of high-stakes testing in relation to teachers' work and identity?

Sub-Questions:

1.1 Perceptions:

- a. What are the perceptions of teachers in relation to the effects of high-stakes tests?
- b. Do teacher perceptions of high-stakes tests differ by independent variables (setting, grade, experience and school performance)?

1.2 Actions:

- a. What actions relating to pedagogy and practice are teachers taking (preparation, time and mode)?
- b. Do actions of preparation, time and mode differ by independent variables (setting, grade, experience and school performance)?

A representative sample of Eastern Tennessee teachers invited to participate in this research study were asked to respond to an online survey instrument. The survey instrument utilized for this study was taken from a nationally normed and validated survey asking teachers to respond to 63 Likert-type questions and 1 open ended question survey format. Data were collected from May 15, 2008 through to June 30, 2008. Respondents representing 6 Eastern Tennessee school systems (n= 408) answered survey questions which addressed teachers' experiences and perceptions working within the high-stakes testing environment in terms of these domains:

- impact on school climate
- pressure experienced
- alignment of pedagogy and practice
- value of test

Upon closure of the survey link the researcher conducted a thorough examination of all data to ensure respondent validity and reliability.

In this chapter I will first review the validity and reliability of the scales and descriptive statistics to provide a demographic exploratory analysis of the sample population and offer a basic description of the study respondents. Then correlational analysis will be presented to describe the strength and the direction of the relationship between the mean scores to the research questions. Additionally, variance found between the mean scores to the specific variables will be reported as a result of correlational tests. Finally, findings from the survey juxtaposed with corroborative narrative data have been

presented in such a way to reflect the significance and strength of the teacher voice. Responses from survey instrument and the analyses of the open-ended question provide an additional lens for establishing reliability of the representative comments of the teacher voice, further validating the data and the findings by the utilization of two data collection methods. Taken together, these data illustrate seminal issues and concerns of teachers that are central to the success of the current high-stakes accountability reform.

Overview of the Sample

The respondents in this sample are teachers who work in Eastern Tennessee school systems. The survey link was accessed a total of 813 times. For analysis, the researcher chose those survey responses which were complete through the primary scale measuring perceptions of high-stakes testing. Of these, a sample of 408 respondents resulted.

Demographic Frequencies

Data were collected and frequencies were run on respondent demographic information as it pertained to survey items associated with:

- grade level
- rural, suburban, urban
- years of teaching experience
- school designation
- Title 1
- Annual Yearly Progress

Table 4:01: Frequencies for school setting

		Frequency	%	Valid %	Cumulative %
Valid	Urban	105	25.7	25.7	25.7
	Suburban	166	40.7	40.7	66.4
	Rural	137	33.6	33.6	100.0
	Total	408	100.0	100.0	

Data collected on schools within the sample were categorized as rural, suburban and urban school settings based on demographically identifying items within the survey instrument. Frequencies show (Table 4:01) that the majority of respondents to the survey instrument were from suburban school settings with 166 respondents, rural school settings with 137 respondents and urban school setting with 105 respondents.

Teachers were presented grade choices grades 3 – 10 and other and asked to identify which grade(s) they taught (Table 4:02). The choice of *other* represented grade groupings such as: K-2, 11-12, multiple grade categories and special support areas spanning several grades. Some teachers answering to this question selected more than one grade designation; therefore, these percentages do not add up to 100.

Examining the grade frequencies, teachers were classified into groupings (Table 4:03) such as: elementary, middle and high school based on the primary grade designation reported. For those few responses where a primary designation could not be clearly determined, the researcher chose to not include these data in the grade level comparisons. Final groupings were 217 elementary teachers and 165 middle school teachers. Those teachers that could not be designated or were in high school were not

Table 4:02: Frequencies for grade

	No		Yes	
	Count	%	Count	%
3	347	85.0 %	61	15.0 %
4	333	81.6 %	75	18.4 %
5	325	79.7 %	83	20.3 %
6	332	81.4 %	76	18.6 %
7	340	83.3 %	68	16.7 %
8	329	80.6 %	79	19.4 %
9	405	99.3 %	3	.7 %
10	402	98.5 %	6	1.5 %
other	355	87.0 %	53	13.0 %

Table 4:03: Frequencies of grade level grouping

		Frequency	%	Valid %	Cumulative %
Valid	Elementary	217	53.2	56.8	56.8
	Middle	165	40.4	43.2	100.0
	Total	382	93.6	100.0	
Missing	High School	10	2.5		
	System	16	3.9		
	Total	26	6.4		
Total		408	100.0		

used in analysis when comparing grade level, although these respondent data were used in other relevant and appropriate comparisons.

Teachers' years of experience (Table 4:04) ranged from 1 to 41 years with an average teaching experience of 15.04 years. The researcher was interested in comparing teachers with experience prior to NCLB with those teachers who only had teaching experience post implementation of NCLB. Therefore, data describing experience was split into two groups: those teachers with less than 6 years teaching experience and those teachers with 7 or more years of teaching experience. Table 4:04 shows the frequencies of these two groups of teachers. Results show that more teachers responding to this survey instrument had experience spanning the years before and after NCLB.

Teachers were asked how their school performance on state-mandated tests compared with that of other schools. Data show (Table 4:05) that 39.7 % of teachers reported that their school was currently an above average performing school. In contrast, 17.9 % teachers reported their school was currently a below average performance school.

The data show (Table 4:06) that 41.2 % of the survey respondents reported they

Table 4:04: Frequencies for teaching experience

		Frequency	%	Valid %	Cumulative %
Valid	6 or fewer years	105	25.7	26.9	26.9
	7 or more years	286	70.1	73.1	100.0
	Total	391	95.8	100.0	
Missing	System	17	4.2		
Total		408	100.0		

Table 4:05: Frequencies for school designation

		Frequency	%	Valid %	Cumulative %
Valid	Above average	162	39.7	39.7	39.7
	Average	173	42.4	42.4	82.1
	Below average	73	17.9	17.9	100.0
	Total	408	100.0	100.0	

Table 4:06: Frequencies for Title 1 and AYP

	Yes	No
Title 1	168 (41.2 %)	240 (58.8 %)
AYP	358 (87.7 %)	50 (12.3 %)

teach in a Title 1 school. In addition, 87.7 % of survey respondents reported that their school is currently making Adequate Yearly Progress (AYP).

High-Stakes Tests and Practice

A frequency was run to determine (Table 4:07) how often high-stakes test results influenced teaching practice. A total of 336 teachers responded to this question with 171 teachers, or 41.9 %, reporting that the high-stakes test influenced their teaching on a daily basis. Additionally, 13.7 % of teachers reported the state-mandated test influenced their teaching a few times a week. Results show that the state-mandated test influences teachers' pedagogy and practice daily and throughout the week for 55.6 % of teachers responding.

Table 4:07: State-mandated test influence on teaching

		Frequency	%	Valid %	Cumulative %
Valid	Daily	171	41.9	50.9	50.9
	A few times a week	56	13.7	16.7	67.6
	A few times a month	51	12.5	15.2	82.7
	A few times a year	38	9.3	11.3	94.0
	Never	20	4.9	6.0	100.0
	Total	336	82.4	100.0	
Missing	I did not receive the school's test results in time to use them	38	9.3		
	I teach a grade and/or subject that does not receive the school's test results	18	4.4		
	I teach a grade and/or subject that should get results but did not receive them	2	.5		
	System	14	3.4		
	Total	72	17.6		
Total		408	100.0		

High-Stakes Testing and Hours of Preparation

Frequencies were run (Table 4:08) to determine how many class hours per year teachers devoted to teaching test-taking skills and preparing students specifically for the state-mandated test. A total of 408 teachers responded to this question; 198 or 48.5 % of the teachers reported spending more than 30 hours in test preparation activities.

High-Stakes Testing and Frequency of Preparation

Frequencies were run (Table 4:09) to determine when teachers practiced test preparation for the state-mandated test. Of the 408 teachers responding to this question,

Table 4:08: Class hours per year spent preparing students for the state-mandated test

		Frequency	%	Valid %	Cumulative %
Valid	None	12	2.9	3.0	3.0
	1-10	64	15.7	16.1	19.1
	11-20	66	16.2	16.6	35.7
	21-30	58	14.2	14.6	50.3
	More than 30	198	48.5	49.7	100.0
	Total	398	97.5	100.0	
Missing	System	10	2.5		
Total		408	100.0		

Table 4:09: Test preparation specifically for the state-mandated test during school year

		Frequency	%	Valid %	Cumulative %
Valid	No specific preparation	12	2.9	3.0	3.0
	Throughout the week before	16	3.9	4.0	7.0
	Throughout the two weeks before	45	11.0	11.3	18.3
	Throughout the month before	56	13.7	14.1	32.4
	Throughout the year	269	65.9	67.6	100.0
	Total	398	97.5	100.0	
Missing	System	10	2.5		
Total		408	100.0		

269 teachers, or 65.9 %, reported that they engaged in test preparation throughout the teaching year.

High-Stakes Tests and Alignment

Frequencies were run (Table 4:10) to determine how similar test preparation materials teachers used in their teaching were to the content of the state-mandated test. Of the 408 teachers responding to this question, 143 teachers, or 35 % of teachers reported they used test preparation materials which were very similar to the content of the state-mandated test. In addition, 224 teachers, or 54.9 %, reported they used test preparation materials which were somewhat similar to the test. Therefore, these two groups of teachers represent 89.9 % of teachers reporting that they do indeed use materials which they report are aligned to the state-mandated test.

Reliability and Validity of Data

The Pedulla et al. (2003) reported high reliability scores utilizing Cronbach's alpha reliability measure to indicate the reliability of a scale. Conceptually and theoretically, Cronbach's alpha indicates how homogeneous items are in a specific scale measuring a single factor.

Initially, the researcher ran factor analysis on all the items from the scales. When calculating the factors from the subscales negative items were reverse coded in order to align all items in the same direction. It is noted that the original scales contained questions which were not used in the data analyses of this current research. However, as

Table 4:10: Alignment of test preparation materials to content of state-mandated test

		Frequency	%	Valid %	Cumulative %
Valid	Very similar	143	35.0	36.0	36.0
	Somewhat similar	224	54.9	56.4	92.4
	Somewhat dissimilar	19	4.7	4.8	97.2
	Very dissimilar	11	2.7	2.8	100.0
	Total	397	97.3	100.0	
Missing	System	11	2.7		
Total		408	100.0		

it was essential to not inhibit or alter the firmly established internal consistency reliability and construct validity of the original scales, all scale questions were surveyed as intact and complete scales (Litwin, 1995). During data analysis a close examination of all scale questions resulted in the researcher omitting six questions which related to ESL, minorities, computers and commercial testing products (see Appendix C). These questions were identified and labeled as distracters and did not serve to support the central research questions of this study:

- What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?
- What are the consequential effects of high-stakes testing in relation to teachers' work and identity?

The purpose of utilizing factor analysis is to measure subscales to ensure that factors do indeed correlate to a specific load on a discrete subscale. Factor analysis determines common variance between each of the variables across all of the subscales.

Mathematically and theoretically, factor analysis groups factors within the conceptual design of the study thereby creating factor groupings which are highly correlated to one another and show little correlation between the factors.

To further address the research questions, a second factor analysis was run on these data. The researcher, finding that some questions loaded across several scales, closely examined each scale question to determine the intent of the question and identify where each question loaded to the scales. All scale questions were coded based on their language construct and intent. Each question was categorized and checked back as to how these respondents were answering. This analysis further contributed to the strength of the findings and ensured the conceptual loading of each question was consistent with the mathematical loading. Based on how items loaded to the factors the researcher removed two questions relating to *cheating* and *fad*. Both questions wanted to load onto scales which did not make sense conceptually. These questions were removed from the data set for conceptual reasons. *Fad* and *cheating* were multiple loaded questions – in close examination of the data these items loaded weakly across three factors. This may have occurred because *fad* is an experientially loaded word with multiple meanings. On close examination of the specific responses to survey items, teachers appeared to be interpreting and determining meaning based on highly individualized experience. *Cheating* showed very little variability across the sample; this makes sense in that there are many structures and procedures in place in the test taking setting to define the physical taking of a high-stakes test. Whereas, *cheating* may be a concern in other

settings, *cheating* showed little variability in these findings. While the purpose of factor analysis is to come up with unique subscales so that each question attributes to a single scale, these two questions were deemed weak and ambiguous in respondents' interpretations. For example, if an ambiguous factor attributed to more than one scale, then that question would influence multiple scales and the result would be an item which is not unique to a specific subscale, thereby attributing little significance to the overall factoring of the scale questions. Therefore, rigorous analysis of the scale questions resulted in the deletion of eight questions from the original survey instrument. After close examination of the respondent data and loadings, the researcher was confident that each of the resulting factors represented logical and discrete variables which factored together in highly characteristic and relational ways (Salkind, 2004). Additionally, a Varimax rotation was employed to ensure that all possible factors loaded to a single variable and to identify the common constructs underling each variable (see Appendix D). As a result, the factor analysis of these data has been established at 4 factors which will be individually discussed:

1. value
2. pressure
3. climate
4. alignment

Factor 1: Value

The strongest factor was determined to be value (Table 4:11). This factor included 14 items such as benefits of the test, test motivation, test validity and reliability and

Table 4.11: Factor loadings for value

Question	Loading
Scores on the state-mandated test accurately reflect the quality of education students have received.	.654
'Overall, the benefits of the state-mandated testing program are worth the investment of time and money.'	.581
The state-mandated test is as accurate a measure of student achievement as a teacher's judgment.	.577
Media coverage of state-mandated test results accurately reflects the quality of education in my state.	.553
The state-mandated test motivates previously unmotivated students to learn.	.542
The state-mandated test has brought much needed attention to education issues in my district.	.499
Differences among schools on the state-mandated test are more a reflection of students' background characteristics than of school effectiveness.	-.475
The state-mandated test measures high standards of achievement.	.467
Media coverage of state-mandated testing issues adequately reflects the complexity of teaching.	.450
State-mandated test results have led to many students being retained in grade in my district.	.428
Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness.	-.396
Media coverage of state-mandated testing issues has been unfair to teachers.	-.374
'If I teach to the state standards or frameworks, students will do well on the state-mandated test.'	.362
Many low scoring students will do better on the state-mandated test if they receive specific preparation for it.	.358

teacher valuing of the test. These examples are descriptive of the scale questions which attributed to this factor and indicate what the factor referred to as value is measuring.

Factor 2: Pressure

The second factor identified was pressure (Table 4:12); this factor represents pressures related to both students and teachers as a result of high-stakes testing. The term used to reference this factor was pressure. This factor examined teachers' responses to questions such as: teacher pressure and performance expectations of students on tests, building administrator's pressure to raise test scores, teaching to the test, curriculum narrowing and teachers identifying their desire to transfer grade. Pressure, an effect of high-stakes testing, accounted for 12 items with a factor load greater than 3.36.

Factor 3: Climate

The third factor determined by the factor analysis was the effect of high-stakes testing on school climate (Table 4:13). This factor of school climate is comprised of items such as: atmosphere for learning, student and teacher morale and high expectations of test performance. Therefore, an additional effect of high-stakes testing is a change to school climate.

Table 4:12: Factor loadings for pressure

Question	Loading
Students are under intense pressure to perform well on the state-mandated test.	.670
Teachers feel pressure from the building principal to raise scores on the state-mandated test.	.570
Many students are extremely anxious about taking the state-mandated test.	.561
There is so much pressure for high scores on the state-mandated test that teachers have little time to teach anything not on the test.	.553
Teachers in my school want to transfer out of the grades where the state-mandated test is administered.	.523
State-mandated testing has caused many students in my district to drop out of high school.	.514
Teachers feel pressure from the district superintendent to raise scores on the state-mandated test.	.494
Teachers feel pressure from parents to raise scores on the state-mandated test.	.492
The state-mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good educational practice.	.485
Administrators in my school believe students' state-mandated test scores reflect the quality of teachers' instruction.	.435
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	.371
'Many students in my class feel that, no matter how hard they try, they will still do poorly on the state-mandated test.'	.336

Table 4:13: Factor loadings for climate

Question	Loading
My school has an atmosphere conducive to learning.	.780
Teachers have high expectations for the in-class academic performance of students in my school.	.654
Student morale is high in my school.	.634
Teacher morale is high in my school.	.611
Teachers have high expectations for the performance of all students on the state-mandated test.	.556
The majority of my students try their best on the state-mandated test.	.431

Factor 4: Alignment

The final factor extracted from the data was alignment of instructional practices and curriculum with the high-stakes tests (Table 4:14). This factor of school alignment is comprised of items such as: test compatibility to daily instruction, curricular alignment, and classroom test alignment to state-mandated test. Therefore, an effect of high-stakes testing is alignment of instructional practices to the test.

Reliability of each of the 4 subscales was run using Cronbach's coefficient alpha. Table 4:15 presents the results of Cronbach's coefficient alpha for each of the subscales. Coefficient alpha measures internal consistency reliability among a group of items to determine how well the different questions work together to measure a single variable (Huck, 2008). Alpha scores exceed .70 with the exception of the weakest score of alignment being .656. Therefore, interpretation of alignment should be noted with

Table 4.14: Factor loadings for alignment

Question	Loading
The state-mandated test is compatible with my daily instruction.	.689
My tests have the same content as the state-mandated test.	.642
My district's curriculum is aligned with the state-mandated testing program.	.574
The state-mandated test is based on a curriculum framework that ALL teachers in my state should follow.	.568
The instructional texts and materials that the district requires me to use are compatible with the state-mandated test.	.489
My tests are in the same format as the state-mandated test.	.412

Table 4.15: Reliabilities

	Cronbach's Alpha
Climate	.709
Pressure	.776
Alignment	.656
Value	.754

caution. Here the lower score describes a level of reliability where teachers responding to the same survey question may respond differently dependent upon intervening variables. For example, if nothing had changed to alter teachers' experience or perceptions, teachers may respond and interpret the question the same or slightly differently.

Results of Primary Research Question

The primary research question asked: What are the consequential effects of high-stakes testing on teachers' pedagogy and practice? To answer the primary research question the data analysis first examined teacher perceptions and how they differed by specific demographics. This analysis was followed by an examination of teacher's actions in relation to the consequential effects of high-stakes testing and how they also differed by specific demographics. Therefore, the sub-questions to the primary research question asked:

- 1.1a. What are the perceptions of teachers in relation to the effects of high-stakes tests?
- 1.1 b. Do teacher perceptions of high-stakes tests differ by independent variables?
- 1.2 a. What actions relating to pedagogy and practice are teachers taking?
- 1.2 b. Do actions of preparation, time and mode differ by independent variables?

Table 4:16: Impact on perception

	Mean	Standard Deviation
Climate	2.97	.346
Pressure	2.94	.370
Alignment	2.78	.446
Value	2.20	.429

High-Stakes Testing: Impact on Perceptions

A first sub-question was formulated and asked: What are the perceptions of teachers in relation to the effects of high-stakes tests (Table 4:16)? To answer this question means were calculated for the 4 scales.

The individual questions were responded to utilizing a 5 point Likert scale, where 1 represents strongly disagree, 3 is neutral and 5 is strongly agree. Comparing the means to the 5 point scale, pressure and climate are neutral because they are very close to 3.0 . Teachers responding to questions about test alignment averaged to be slightly negative at 2.7. These findings show that respondents are slightly disagreeing that they are aligning their practice and pedagogy with the state-mandated test. Value is a disagree at 2.20, which shows that respondents on average do not find value in the state-mandated test.

Perception Difference by Independent Variables

A second sub-question was formulated and asked: Do teacher perceptions of high-stakes tests differ by independent variables? To answer this sub-question MANOVA's

Table 4:17: Perception difference by grade

Dependent Variable	grade	Mean	P value
Value	Elementary	2.204	.444
	Middle	2.176	
Pressure	Elementary	2.987	.002
	Middle	2.868	
Climate	Elementary	3.032	.009
	Middle	2.912	
Align	Elementary	2.783	.878
	Middle	2.776	

were run looking for differences in all 4 subscales by grade, performance, setting and experience to determine how teachers' perceptions differed by demographic.

Perception Difference by Grade

A MANOVA was run to examine all 4 subscales at once to determine if there were any differences by grade (Table 4:17). The results of the MANOVA, utilizing Wilks' Lambda Multivariate test for grade, are $F(4,377)=5.704$, $p<.001$ which indicates that at least one scale differs. Individual ANOVA's were run to determine which of the scales differed. Significant differences were found with pressure ($p=.002$) and climate ($p=.009$). No differences were found with alignment ($p=.878$) or value ($p=.444$). For pressure, the mean for elementary school teachers was 2.99 and for middle school teachers the mean was 2.87. For climate, the mean for elementary school teachers was 3.03 and for middle school teachers the mean was 2.91. Therefore, because there are only two groups, we can examine the means and determine that for both pressure and climate elementary school teachers have significantly higher means than middle school teachers.

In both instances, elementary school teachers perceived greater impact on both scales of pressure and climate than middle school teachers. Therefore, elementary school teachers feel more pressure and impact on climate than middle school teachers.

Perception Difference by Setting

A MANOVA was run to examine all 4 subscales at once to determine if there were any differences by setting (Table 4:18). The results of the MANOVA, utilizing Wilks' Lambda Multivariate test, are $F(8,804)=2.638$, $p=.007$ which indicates that at least one scale differs by setting. Individual ANOVA's were run to determine which of the scales differed. Significant differences were found with value ($p=.037$) and climate ($p=.001$). No differences were found with alignment ($p=.915$) and pressure ($p=.180$). For value, the mean for urban schools was 2.13, suburban schools was 2.21 and rural schools was 2.24. Because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. Suburban schools do not differ from urban schools ($p=.111$) or from rural schools ($p=.806$). However, rural schools and urban schools do differ ($p=.035$). Urban schools value the state-mandated test a slight amount less than the rural schools. For climate, the mean for urban schools was 2.83, suburban schools was 3.04 and rural schools was 3.00. Pairwise comparisons found that urban schools differ from suburban schools ($p=.001$) and rural schools ($p=.011$). No significant difference was found between suburban schools and rural schools ($p=.742$). As a result, teachers teaching in suburban schools and rural schools feel greater impact on climate than teachers of urban school settings.

Table 4:18: Perception difference by setting

Dependent Variable	School setting	Mean	P value
Value	Urban	2.127	.037
	Suburban	2.213	
	Rural	2.238	
Pressure	Urban	2.989	.180
	Suburban	2.935	
	Rural	2.900	
Climate	Urban	2.833	.001
	Suburban	3.036	
	Rural	2.999	
Align	Urban	2.786	.915
	Suburban	2.769	
	Rural	2.788	

Perception Difference by Experience

Correlations were run to examine if there were relationships between the 4 scales with teaching experience (Table 4:19). A significant weak relationship was found between pressure and experience ($r=-.153$, $p<.001$). Results show that as teaching experience increases, teacher perceptions of pressure tends to decrease. While looking at these groups numerically, the researcher chose to aggregate the data to include looking at teacher experience by those teachers who have experience before the NCLB Act (2002) and those teachers who entered the profession after the mandates of the NCLB Act. Thus, two groups of teachers are defined:

1. those teachers with 7 or more years teaching experience
2. those teachers with 6 or fewer years teaching experience

Table 4:19: Perception difference by experience

Dependent Variable	Years of experience	Mean
Value	6 or fewer years	2.213
	7 or more years	2.184
Pressure	6 or fewer years	2.967
	7 or more years	2.935
Climate	6 or fewer years	2.970
	7 or more years	2.969
Alignment	6 or fewer years	2.792
	7 or more years	2.769

To compare the two groups a MANOVA was run. Results were $F(4,386)=.409$, $p=.803$.

No significant differences were found between these groups. In conclusion, when examining experience numerically, the analysis showed there was a weak relationship with pressure.

Perception Difference by Performance

A MANOVA was run to look at all 4 subscales at once to determine if there were any differences by school performance (Table 4:20). The results of the MANOVA, utilizing Wilks' Lambda Multivariate test, are $F(8,804)=7,920$, $p<.001$ which indicates that at least one scale differs by school performance. Individual ANOVA's were run to determine which of the scales differed. Significant differences were found with value ($p<.001$), pressure ($p=.006$) and climate ($p<.001$). No differences were found with alignment ($p=.704$). For value, the mean for above average performance schools was

2.22, the mean for average performance schools was 2.24, and the mean for below average performance schools was 2.06. Therefore, because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. Findings showed that below average performance schools do differ from average ($p < .001$) and above average performance schools ($p = .002$). However, no differences were found between above average performance schools and average performance schools ($p = .912$). Teachers teaching in schools with below average performance assign a significantly lower value of high-stakes tests than those teachers teaching in schools with average or above average performance. No differences in perception of value were found between teachers teaching in schools with average performance and those with above average performance.

Table 4:20: Perception difference by performance

Dependent Variable	School Performance	Mean	P value
Value	Above average	2.223	<.001
	Average	2.238	
	Below average	2.058	
Pressure	Above average	2.915	.006
	Average	2.906	
	Below average	3.062	
Climate	Above average	3.118	<.001
	Average	2.941	
	Below average	2.717	
Align	Above average	2.763	.704
	Average	2.801	
	Below average	2.767	

For pressure, the mean for above average performance schools was 2.92, for average performance schools was 2.91, for below average performance schools was 3.06. Because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. Below average performance schools do differ from average performance schools ($p < .007$) and above average performance schools ($p = .013$). However, no differences were found between above average performance schools and average performance schools ($p = .969$). Teachers teaching in schools with below average performance perceive higher levels of pressure than those teachers teaching in schools with average or above average performance. No differences in perception of pressure were found between teachers teaching in schools with average performance and those with above average performance.

For climate, the mean for schools with above average performance was 3.12, for schools with average performance was 2.94, for schools with below average performance was 2.72. Therefore, because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. All three performance groups differed significantly from one another with all p-values less than .001. Teachers teaching in schools with below average performance have the largest mean and those reporting from above average performance schools have the lowest mean. Therefore, as school test performance decreases, impact on school climate increases. As school performance decreases value of the test decreases, while pressure and impact on school climate tends to increase.

High-Stakes Testing: Impact on Pedagogy and Practice

Teachers reported several ways in which high-stakes tests impacted their pedagogy and practice. Table 4:21 describes the frequency of teachers who indicated impact in pedagogy and practice. The data show that 90.7 % of the respondents teach test-taking skills, with 93.6 % of teachers reporting that they encourage their students to work hard and prepare for the test. Only 21.1 % of teachers provided rewards for test completion. Teaching to the standards or frameworks known to be on the test was a regular practice of 90.7 % of the responding teachers. Providing students with items similar to those on the test was the practice of 79.4 % teachers. Additionally, 72.3 % of the respondents reported they provide test-specific preparation materials developed

Table 4:21: Preparation: impact on pedagogy and practice

	Count	%
I teach test-taking skills	370	90.7 %
I encourage students to work hard and prepare	382	93.6 %
I provide rewards for test completion	86	21.1 %
I teach the standards or frameworks known to be on the test	370	90.7 %
I provide students with items similar to those on the test	324	79.4 %
I provide test-specific preparation materials developed commercially or by the state	295	72.3 %
I provide students with released items from the state-mandated test	176	43.1 %

commercially or by the state, as well 43.1 % reported that they provided students with released items from the state-mandated test.

Impact of Preparation by Grade

Chi-Square tests were run to measure the relationship of grade with areas of preparation. The researcher was interested if preparation differed by grade. Results of the Chi-Square tests are in Table 4:22. A Pearson Chi-Square guided the interpretation of these results primarily to evaluate the relationship by assessing the significant difference between the expected and actual frequencies in each of the variables - grade and preparation (Nardi, 2006b).

Table 4:22: Impact of preparation by grade

	Elementary		Middle		p-value
	Count	%	Count	%	
I teach test-taking skills	208	95.9 %	141	85.5 %	<.001
I encourage students to work hard and prepare	209	96.3 %	152	92.1 %	.075
I provide rewards for test completion	50	23.0 %	31	18.8 %	.314
I teach the standards or frameworks known to be on the test	202	93.1 %	147	89.1 %	.168
I provide students with items similar to those on the test	180	82.9 %	124	75.2 %	.061
I provide test-specific preparation materials developed commercially or by the state	166	76.5 %	114	69.1 %	.105
I provide students with released items from the state-mandated test	104	47.9 %	62	37.6 %	.043

Overall, 91 % of teachers teach test-taking skills. More specifically, 95.9 % of elementary teachers teach test-taking skills and 85.5 % of middle schools teachers teach test-taking skills. The results of the Chi-Square were $\chi^2(1)=12.84, p<.001$. Differences in these two variables demonstrate statistically significant differences between grade and test taking skills. Results indicated that elementary teachers are significantly more likely to teach test-taking skills than middle school teachers.

Within grade variance showed that 47.9 % of elementary school teachers and 37.6 % of middle schools teachers do provide released items from the state-mandated test. The results of the Chi-Square were $\chi^2(1)=4.09, p=.048$. Therefore, elementary school teachers are significantly more likely to provide released test items to their students for test preparation.

There was a marginal difference in both elementary (82.9 %) and middle school teachers (75.2 %) providing students with items similar to those items on the test. The results of the Chi-Square were $\chi^2(1)=3.51, p = .073$. Although not significant at $p = .05$, it may be considered marginal since it is less than .10. Therefore, there is a possibility that elementary teachers may provide similar test items more frequently than middle school teachers.

No significant difference in elementary or middle school teachers were found with encouraging students to work hard and prepare for the high-stakes test ($p=.111$). Additionally, there was no statistical significance in how elementary or middle school teachers provided rewards for test completion, ($p= .314$.) No difference was found in how

Table 4:23: Impact of preparation by setting

	Urban		Suburban		Rural		p-value
	Count	%	Count	%	Count	%	
I teach test-taking skills	92	87.6 %	155	93.4 %	123	89.8 %	.256
I encourage students to work hard and prepare	98	93.3 %	156	94.0 %	128	93.4 %	.971
I provide rewards for test completion	31	29.5 %	24	14.5 %	31	22.6 %	.011
I teach the standards or frameworks known to be on the test	90	85.7 %	157	94.6 %	123	89.8 %	.045
I provide students with items similar to those on the test	76	72.4 %	134	80.7 %	114	83.2 %	.102
I provide test-specific preparation materials developed commercially or by the state	74	70.5 %	121	72.9 %	100	73.0 %	.889
I provide students with released items from the state-mandated test	47	44.8 %	75	45.2 %	54	39.4 %	.557

both elementary teachers or middle school teachers teach to the standards ($p = .199$). And, no difference was found between elementary and middle school teachers in frequency of provision of test-specific preparation materials, ($p = .129$).

Impact of Preparation by Setting

Chi-Square tests were run to measure the relationship of setting with areas of preparation. The researcher was interested if preparation differed by setting. Results of the Chi-Square tests are in Table 4:23.

The results of the Chi-Square between setting and rewards were $\chi^2(2) = 9.07$, $p = .011$. There was a significant difference between urban, suburban and rural teachers providing rewards for test completion. The results show that urban school teachers (29.5

%) are more likely to provide rewards for test completion, suburban school teachers (14.5%) are less likely to provide rewards, with rural school teachers (22.6 %) falling in the middle. Therefore, urban school teachers are more likely to provide rewards for test completion and suburban school teachers are less likely.

The results of the Chi-Square between setting, frameworks and standards were $\chi^2(2)=6.18, p = .045$. Teaching the standards or frameworks known to be on the test results showed that significant differences existed. Again, rural school teachers at 89.8 % are falling in the middle, which is equal to the overall of 90.7 %. Urban school teachers at 85.7 % are less likely to teach to the test, whereas, suburban school teachers at 94.6 % are more likely to teach to the standards or frameworks known to be on the test.

There was no difference found in the teaching of test-taking skills and encouraging students to work hard and prepare for state-mandated tests. ($p=.971$). No differences were found with providing students with items similar to those on the test, $p=.102$. No differences were found with teachers providing test-specific preparation materials developed commercially or by the state, $p=.889$. No differences were found with use of released items from the state-mandated test, $p=.557$.

In sum, significant differences were found with rural school teachers more likely to provide rewards for test completion and suburban school teachers were more likely to teach to the standards or frameworks known to be on the test. No other setting differences were found in preparation.

Impact of Preparation by Experience

Analysis looked at whether preparation differed by years of teaching experience. The analysis compared teacher preparation with two groups of teachers: those teachers with teaching experience before testing reforms of the 2002 NCLB Act and those teachers who had teaching experience only in those years following the implementation of state-mandated testing reforms of the NCLB Act. Chi-Square tests were run to test for differences between preparation and years of teaching experience. No differences were found between experience and any areas of preparation.

Impact of Preparation by Performance

Chi-Square tests were run to look for differences between preparation and performance. Results of the Chi-Square tests are in Table 4:24. When significant relationships were found, the adjusted residual was used as an indicator of where differences occurred. When the residual was larger than a magnitude of 2 this indicated that there was something different occurring other than the overall expected.

The results of the Chi-Square between performance and rewards were $\chi^2(2)=6.38$, $p=.041$). Therefore, significant differences were found with teachers providing rewards for test completion. Teachers teaching in above average performing schools (14.8 %) are less likely to provide rewards; whereas, both groups of teachers teaching in average performing schools (24.9 %) and those teachers teaching in below average performing schools (26.0 %) are more likely to provide rewards for test completion.

Table 4:24: Impact of preparation by performance

	Above Average		Average		Below Average		p-value
	Count	%	Count	%	Count	%	
I teach test-taking skills	148	91.4 %	155	89.6 %	67	91.8 %	.805
I encourage students to work hard and prepare	154	95.1 %	160	92.5 %	68	93.2 %	.617
I provide rewards for test completion	24	14.8 %	43	24.9 %	19	26.0 %	.041
I teach the standards or frameworks known to be on the test	149	92.0 %	159	91.9 %	62	84.9 %	.175
I provide students with items similar to those on the test	125	77.2 %	142	82.1 %	57	78.1 %	.513
I provide test-specific preparation materials developed commercially or by the state	119	73.5 %	125	72.3 %	51	69.9 %	.850
I provide students with released items from the state-mandated test	69	42.6 %	76	43.9 %	31	42.5 %	.962

No differences were found with teaching to the standards or frameworks on the test ($p=.175$), with providing students items similar to those on the test ($p=.513$), with providing test-specific preparation materials developed commercially or by the state ($p=.850$), with providing students with released items from the state-mandated test ($p=.962$), with teachers teaching test-taking skills ($p=.805$) and with teachers encouraging students to work hard and prepare for state-mandated tests ($p=.617$).

Descriptive Statistics—Instructional Time

Teachers were asked how time spent on specific strategies and activities has been affected due to state-mandated testing. In order to determine the areas showing the most

change descriptive statistics were run. The mean and standard deviation were computed for each of the questions (Table 4:25). Within the scale any score above a 3 will indicate an increase and any score below a 3 will indicate a decrease. For example, a mean of 4.38 shows that teachers are moderately increasing instruction in tested area.

Teachers report they are increasing instruction in tested areas with high-stakes attached, with a mean score of 3.80. As well, a mean of 3.38 indicated that teachers are increasing the amount of parental contact as a result of high-stakes tests. All other areas such as class trips, field trips, free time, enrichment activities, instruction in the fine arts and other curricular related areas have decreased as a result of high-stakes tests.

Instructional Time by Grade

A MANOVA was run to determine if there were any differences by grade (Table 4:26). The results of the MANOVA, utilizing Wilks' Lambda Multivariate test for grade, are $F(16,344)=3.155$, $p<.001$ which indicates there are areas of difference. Individual ANOVAs were run to determine which of the scales differed. Significant differences were found with instruction in tested areas ($p=.013$) and instruction in areas not covered by the state-mandated test ($p=.043$). For instruction in tested areas, the mean was 4.47 for elementary school teachers and 4.28 for middle school teachers. For instruction in areas not covered by the state-mandated test, the mean for elementary teachers was 1.77 and for middle school teachers the mean was 1.97. Therefore, because there are only two groups, we can look at the means and determine that for instruction in tested areas and in areas not covered by the state-mandated test we can determine that these two areas are most affected by state-mandated tests. Elementary teachers are increasing their

Table 4:25: Instructional time

	N	Minimum	Maximum	Mean	Std. Deviation
Instruction in tested areas	387	3	5	4.38	.733
Instruction in tested areas with high stakes attached (e.g., promotion, graduation, teacher rewards)	387	1	5	3.80	.886
Parental contact	386	1	5	3.38	.857
Instruction in tested areas without high stakes attached	387	1	5	2.77	1.064
Instruction in physical education	387	1	5	2.75	.835
Administrative school assemblies (e.g., award ceremonies)	386	1	5	2.52	.835
Instruction in the fine arts	387	1	5	2.51	.877
Classroom enrichment activities (e.g., guest speakers)	386	1	5	2.38	.979
Instruction in industrial/vocational education	387	1	5	2.37	.897
Student choice time (e.g., professional choral group performances)	386	1	5	2.35	.856
Instruction in foreign language	387	1	5	2.33	.937
Student free time (e.g., recess, lunch)	386	1	5	2.28	.847
Student performance (e.g., class plays)	386	1	5	2.25	.910
Field trips (e.g., museum tour, hospital tour)	386	1	5	2.12	.912
Class trips (e.g., circus, amusement park)	386	1	5	1.92	.934
Instruction in areas not covered by the state-mandated test	387	1	5	1.87	.957
Valid N (listwise)	386				

instruction to tested areas more than middle school teachers. The instruction in areas not covered by the state-mandated test is decreasing more in elementary schools than in the middle schools. Therefore, it can be determined that state-mandated tests have greater effect at the elementary school level than at the middle school level. No other differences were found.

Instructional Time by Setting

A MANOVA was run to determine if there were differences by setting. Looking at how teachers reported their instructional time and if it was impacted in relation to the state-mandated test determined no difference between settings. Results indicated $F(32,736)=1.015$, $p=.446$, no significant differences were found between settings for change resulting from the state-mandated test.

Instructional Time by Experience

A MANOVA was run to determine if there were differences by experience. Looking at how teachers reported their instructional time and if it was impacted in relation to the state-mandated test determined no difference between experience. Results indicated $F(16,362)=1.090$, $p=.362$, no significant differences were found between experience for change resulting from the state-mandated test.

Instructional Time by Performance

A MANOVA was run to look if there were any differences in instructional time by performance (Table 4:27). The results of the MANOVA, utilizing Wilks' Lambda Multivariate test, are $F(32,736)=1.865$, $p=.003$ which indicates differences in instructional time by performance. Individual ANOVA's were run to determine which of

Table 4:26: Instructional time by grade

Dependent Variable	grade	Mean	p-value
Instruction in tested areas	Elementary	4.468	.013
	Middle	4.276	
Instruction in areas not covered by the state-mandated test	Elementary	1.771	.043
	Middle	1.974	
Instruction in tested areas with high stakes attached (e.g., promotion, graduation, teacher rewards)	Elementary	3.771	.426
	Middle	3.846	
Instruction in tested areas without high stakes attached	Elementary	2.766	.800
	Middle	2.737	
Instruction in the fine arts	Elementary	2.502	.753
	Middle	2.532	
Instruction in physical education	Elementary	2.688	.183
	Middle	2.808	
Instruction in foreign language	Elementary	2.341	.395
	Middle	2.256	
Instruction in industrial/vocational education	Elementary	2.361	.930
	Middle	2.353	
Student free time (e.g., recess, lunch)	Elementary	2.229	.172
	Middle	2.353	
Field trips (e.g., museum tour, hospital tour)	Elementary	2.171	.273
	Middle	2.064	
Class trips (e.g., circus, amusement park)	Elementary	1.854	.201
	Middle	1.981	
Student choice time (e.g., professional choral group performances)	Elementary	2.298	.168
	Middle	2.423	
Administrative school assemblies (e.g., award ceremonies)	Elementary	2.439	.094
	Middle	2.590	
Classroom enrichment activities (e.g., guest speakers)	Elementary	2.415	.308
	Middle	2.308	
Student performance (e.g., class plays)	Elementary	2.151	.059
	Middle	2.333	
Parental contact	Elementary	3.341	.305
	Middle	3.436	

Table 4:27: Instructional time by performance

Dependent Variable	School performance	Mean	p-value
Instruction in tested areas	Above average	4.377	.747
	Average	4.370	
	Below average	4.448	
Instruction in areas not covered by the state-mandated test	Above average	1.896	.084
	Average	1.945	
	Below average	1.642	
Instruction in tested areas with high stakes (e.g., promotion, rewards)	Above average	3.825	.191
	Average	3.715	
	Below average	3.940	
Instruction in tested areas without high stakes attached	Above average	2.753	.018
	Average	2.897	
	Below average	2.463	
Instruction in the fine arts	Above average	2.630	.089
	Average	2.455	
	Below average	2.388	
Instruction in physical education	Above average	2.864	.040
	Average	2.709	
	Below average	2.567	
Instruction in foreign language	Above average	2.403	.148
	Average	2.327	
	Below average	2.134	
Instruction in industrial/vocational education	Above average	2.344	.313
	Average	2.442	
	Below average	2.254	
Student free time (e.g., recess, lunch)	Above average	2.383	.163
	Average	2.218	
	Below average	2.209	
Field trips (e.g., museum tour, hospital tour)	Above average	2.162	.248
	Average	2.158	
	Below average	1.955	
Class trips (e.g., circus, amusement park)	Above average	1.994	.325
	Average	1.909	
	Below average	1.791	
Student choice time (e.g., professional choral performances)	Above average	2.487	.003
	Average	2.352	
	Below average	2.060	
Administrative school assemblies (e.g., award ceremonies)	Above average	2.494	.109
	Average	2.606	
	Below average	2.358	

Table 4.27, cont.

Dependent Variable	School performance	Mean	p-value
Classroom enrichment activities (e.g., guest speakers)	Above average	2.383	.662
	Average	2.412	
	Below average	2.284	
Student performance (e.g., class plays)	Above average	2.318	.005
	Average	2.321	
	Below average	1.925	
Parental contact	Above average	3.377	.990
	Average	3.388	
	Below average	3.373	

the areas of instructional time differed (see Table 4:27). Significant differences were found with instruction in tested areas without high stakes attached ($p=.018$), instruction in physical education ($p=.040$), student choice time ($p=.003$) and student performance ($p=.005$). No differences were found with any other areas of instructional time.

For instruction in tested areas without high stakes attached, (e.g, science and writing) the mean for above average performance schools was 2.75, the mean for average performance schools was 2.90, and the mean for below average performance schools was 2.46. Therefore, because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. Findings showed that below average performance schools do differ from average performance schools ($p=.013$). However, no differences were found between above average performance schools and average performance schools ($p=.445$) and between above average performance schools and below average performance schools ($p=.145$). Therefore, teachers who teach in schools with below average performance tend to spend decreasing time in tested areas without high stakes attached.

For instruction in physical education, the mean for above average performance schools was 2.87, for average performance schools was 2.71, for below average performance schools was 2.57. Because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. Below average performance schools do differ from above average performance schools ($p=.040$). However, no differences were found between average performance schools and below

average performance schools ($p=.467$) and average performance schools and above average performance schools ($p=.222$). Teachers who teach in schools with below average performance tend to spend decreasing instructional time given to physical education.

For student choice time, the mean for schools with above average performance was 2.49, for schools with average performance was 2.35, for schools with below average performance was 2.06. Therefore, because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. Below average performance schools differ from above average performance schools ($p=.002$) and average performance schools ($p=.046$). There is no difference between above average performance schools and average performance schools in relation to time given to student choice ($p=.326$). Therefore, the below average schools have a greater negative impact on student choice time than either the above average performance schools or average performance schools.

For student performance, the mean for schools with above average performance was 2.32, for schools with average performance was 2.32, for schools with below average performance was 1.93. Therefore, because we have more than 2 groups, pairwise or post hoc comparisons were run to determine how the means differ. Below average performance schools differ from above average performance schools ($p=.009$) and average performance schools ($p=.007$). There is no difference between above average performance schools and average performance schools in relation to time given to student

performance ($p=1.00$). Therefore, the below average performance schools have a greater negative impact on student performance time than either the above average performance schools or average performance schools.

Descriptive Statistics—Modes of Instruction

The teachers were asked if state-mandated tests have influenced time spent on modes of instruction. Descriptive statistics were run in order to determine which modes of instruction were influenced the most. The mean and standard deviation were computed for each of the questions (Table 4:28). Within the scale any score above a 3 will indicate a greater influence on modes of instruction and any score below a 3 will indicate a lack of influence on modes of instruction. For example, a mean of 4.50 shows that time spent directed to a specific mode was influenced by state-mandated testing.

Table 4:28: Modes of instruction

	N	Minimum	Maximum	Mean	Std. Deviation
Problems that are likely to appear on the state-mandated test	391	1	4	3.15	.699
Basic skills	391	1	4	3.04	.699
Critical thinking skills	391	1	4	2.97	.717
Concept development using manipulatives or experiments	391	1	4	2.83	.757
Whole group instruction	391	1	4	2.83	.695
Individual seat work	391	1	4	2.77	.697
Students working together in small groups (cooperative learning)	391	1	4	2.76	.774
Valid N (listwise)	391				

Teachers reported that time focused on questions which are likely to appear on the state-mandated test were slightly influenced based on a mean of 3.15. Basic skill had a mean of 3.04 which indicates that teachers neither agreed nor disagreed that instructional time directed to basic skills was influenced by the state-mandated test. All other modes of instruction had means of 3.00 or less which indicates that there was little to no influence on time spent directed toward these modes of instruction.

Modes of Instruction by Grade

A MANOVA was run to determine if there were differences by grade and how teachers reported if instructional time by mode of instruction was impacted in relation to the state-mandated test. Results were $F(7,362)=1.77$, $p=.093$. No significant differences were found between grade for impact from the state-mandated test on modes of instruction.

Modes of Instruction by Setting

A MANOVA was run to determine if there were differences by school setting and how teachers reported if instructional time by mode of instruction was impacted in relation to the state-mandated test. Results were $F(14,762)=1.13$, $p=.326$. No significant differences were found between setting for impact from the state-mandated test on modes of instruction.

Modes of Instruction by Experience

A MANOVA was run to determine if there were differences by years of teaching experience and how teachers reported if instructional time by mode of instruction was impacted in relation to the state-mandated test. Results were $F(7,383)=1.94$, $p=.062$. No

significant differences were found between levels of experience in relation to impact from the state-mandated test on modes of instruction.

Modes of Instruction by Performance

A MANOVA was run to determine if there were differences by school performance and how teachers reported if instructional time by mode of instruction was impacted in relation to the state-mandated test. Results were $F(14,764) = .904$, $p = .554$. No significant differences were found between levels of school performance in relation to impact from the state-mandated test on modes of instruction.

Open-Ended Survey Question: Narrative Data

The final section of the survey instrument asked teachers to optionally respond to a narrative formatted question. The secondary research question asked:

- What are the consequential effects of high-stakes testing in relation to teachers' work and identity?

Data included a representative sample of 112 teachers who provided corroborative narrative data in response to a single question about their work and identity within the high-stakes testing environment.

Utilizing a highly corroborative process, manual coding techniques (Miles & Huberman, 1994; Patton, 1990) and Text Analysis SPSS for Survey Research software the researcher employed an iterative methodology in determining corroborative themes presented in the data. This iterative process used for data analysis provided an authentic, corroborative measure to support the documentation of the findings. These narrative data

were analyzed and triangulated with the quantitative data. As a result, the inclusion of teacher voice to corroborate the findings of the quantitative data further supports the generalization of these findings.

Results of the open-ended data analysis determined five categories which were representative of the broad teacher response to the second research question. While these five categories represent 27.4 % of the total respondent sample, N=408, they embody the range of teacher voice in response to the second research question. These categories were:

1. Test preparation and curriculum narrowing
2. Teacher identity
3. Test use and value of test
4. Sociocultural influences and intervening variables
5. Rewards and sanctions

Teacher responses categorized to these five broad themes stand as significant representational data of the respondent voice. While each respondent entry was determined and categorized to be strongly attributed to the specific theme, each respondent data entry holds characteristics and is closely connected to one or more of the categories.

Test Preparation and Curriculum Narrowing

Teachers expressed concern for the extensive ways in which high-stakes tests affects their work and teacher identity. Several teachers reflected back over their years of teaching making comparative statements describing the increased impact of the current

high-stakes testing environment on their work. Teachers further described their perceptions of themselves as teachers and the role they have taken on as teachers within the high-stakes testing environment. Specifically, these teacher voices spoke of their experience of being influenced, manipulated and threatened to comply and align their teaching pedagogy, practice and identity to the expectations of the current high-stakes test culture. Here, three distinctive teacher voices rise – a rebel, a renegade and a pragmatist.

(Respondent 407) *Sixteen years ago my first group of students were able to take an open-ended assignment and run with it---no hand-holding from me. I gave the assignment structure without limiting the creativity and critical thinking of the students. Fast forward -- the same assignment requires hand-holding from me for MOST of the students in an advanced class. These kids can't think for themselves, but they can bubble in bubble sheets. Since the advent of high-stakes testing in TN, teachers are unable to stretch students' minds and help them think critically about the world around them. We don't have time to show students how everything in this world--the natural, the political, the societal, the human is all inter-related. We are shooting ourselves in the foot with high-stakes testing. Soon, all creative thought, all innovation will be extinct within our society, except in perhaps small pockets here and there. As far as how this relates to my identity as a teacher... for me, I want students to think for themselves, to have a joy of learning, an excitement for learning, to question, to invent, to create and my hands are tied ... unless I want to be a rebel ...and sometimes I am – but not nearly as often as I ought to be. When our society has fallen behind the rest of the world in terms of innovation and problem-solving, we only have ourselves to thank.*

(Respondent 414) *As a teacher, I am aware of the pressures of high stakes testing; however, I choose to use good teaching practices in general rather than test-prep materials to prepare my students for the tests. In doing so, most of my students do well anyway. I despise the amount of attention and importance placed on the testing and results and disagree with how they are used to rate teachers, schools, and districts. I know too many teachers who only teach to the test and find themselves frustrated with what the kids can and can't do because of the test-prep they've*

received in place of real education. It is discouraging to say the least, but I feel like I am a renegade in the sense that I know how to improve my score and more importantly my students' knowledge base--actually teach them content and skills, not just test-prep. I see test-prep as an incredible waste of instructional time. I wish more money was spent on good teaching materials or training better teachers rather than test-prep booklets and programs (i.e. Thinklink). However, because the results are tied to my job, I do encourage students to take the test seriously and to try to do their best. I try to alleviate the stress about the test through encouraging them that they know most of the material and that their effort is what counts the most. Despite what I say or do or what my school says or does, some students still don't take the test seriously and some purposely fill in random answers to punish a subject area teacher they dislike or to punish the school as a whole. This is especially true of 8th grade students who know the test scores or the school's reputation are no longer their problems since they move on to high school anyway.

(Respondent 673) *There is only one reason why teachers in my school do not want to transfer out of the grades where the state-mandated test is administered. I teach at a Title I school where Reading First is wholeheartedly embraced and practiced in grades K-3. Teachers in these grades are required to devote MANY hours to Reading First training. This occurs after school for several hours outside of our normal teaching contract time. Although these teachers receive a monetary stipend for attending, fourth and fifth grade teachers do not want to have to do this. Our current fifth grade students have received Reading First training since K. Veteran classroom teachers have observed that these students are no better off for having had to endure the kind of teaching and learning that are required as elements of Reading First. State-mandated testing IS leading teachers at my school, and teachers are being led to teach in ways that contradict their own ideas of good educational practice. Teachers at my school are fortunate that the fire marshal has required us to stop using door stops to prop our classroom doors open; It is MUCH easier for us to teach in ways that we know are best for kids when our doors are shut. I graduated with a master's degree in Elementary Teaching within the past five years and at the institution where I completed this training, we did not spend much time talking about high-stakes testing and how it would impact our teaching.*

Teacher Identity

Teachers responding to the open-ended question of teacher work and identity expressed feelings of stress and pressure related to high-stakes test mandates. As evidenced in the respondent data, teachers articulated a loss of professionalism and autonomy. Moreover, working in highly scripted and test responsive environments stated that their teacher knowledge and best practice had been positioned to the sidelines of their teaching experience.

For example, these three respondent data sets present teacher voice articulating a loss of respect and how this experience impacted teacher identity. Teachers conveyed their feelings of stress and pressure have become so elevated there is no joy in their work. Finally, teachers communicated their sense of powerlessness as they have been forced to teach a highly scripted program within a work environment where frustration and heartbreak is the result.

(Respondent 306) *In our school district, superintendents and administrators receive bonuses if we do well, but teachers do not. If we do poorly, teachers are blamed above all. I feel that there is little respect for teachers anymore - everyone is sure they could do it better than us. They compliment us and call us the "experts", but they ignore our opinions and treat us like idiots. I admit I don't feel like putting much effort into a job in which I am so little appreciated. I have considered other careers and continue to do so. The new teachers do not seem to have the commitment to their jobs, but I'm beginning to think it's a better attitude considering the climate we teach in. They seem to approach it more as a job and spend less preparation time than we did as new teachers.*

(Respondent 721) *TCAP testing is so focused on and stressed at our school that I feel I don't truly get the chance to ENJOY my students! It aggravates me to be told that MY scores could be better when it's actually*

the students who are making the scores...NOT ME..I have done my job to the best of my ability. Why should I be judged by their scores?

(Respondent 798) *I feel that our county has forced teachers to "teach to the test" by pulling departmentalization out of our upper elementary grades and butchering our one solid curriculum. They have mandated a reading program that completely swallows the majority of our instructional day. Our Social Studies and Science time has plummeted (if we have any at all), and our children are not prepared academically and intellectually as in years past. I feel as though I teach Math and Reading all day long without any opportunity to work on Science and Social Studies even though our state standards in those two subjects are immense! I am beyond frustrated with this new mandated reading series that allows zero flexibility and/or integration with other subjects. The county acts on the surface as though the teachers have choice. WE DO NOT. Our curriculum is dictated by select Central Office staff members, and our children are suffering from a severely flawed system that is preventing them from gaining the necessary knowledge and skills required in elementary school. It is absolutely heartbreaking. I feel as though we are setting up our children for failure rather than preparing them for success. I feel as though I am forced to teach against my principles (teaching to a test) and against the very fabric of my educational philosophy. One individual in power should not be allowed to destroy the curriculum for an entire system due to personal control issues and a power hungry nature. We teachers, are powerless. All we hear is what a "good teacher" is/does Poor Central Office choices do not equate in me being a poor teacher. We are working harder not smarter just so a select few can maintain high paying jobs. Their ineptitude, arrogance, and inflexibility does not mean I am an ineffective teacher, yet Central Office makes us feel that way. EACH AND EVERY DAY.*

Test Use and Value of Test

Teachers reported how the use of high-stakes tests impacted their work and teacher identity. While many teachers confirmed their support of both standards and accountability, teachers were cognizant of a powerful test reliant system which they described as being one reduced to the production of test scores. Teachers responding to the open-ended research question were generally concerned with the validity and use of a

single test score. These teachers stated they felt caught within a system which judged them but did not include them.

(Respondent 249) *I am concerned about the how the tests are used. I feel that the scores are used to make assumptions about students, and schools that are incomplete. I have no problems in using the data gathered by such tests. I know from years of experience, however, that some, students do better than they should (guessing) and some, students experience test anxiety (thus not showing what they are capable of doing.)*

(Respondent 471) *It is very inaccurate to compare schools within a certain county/state because the schools do not have a consistent way of testing. For example, my school has a very strict, consistent method. We switch grade levels to test, we never have our own test alone, etc. Other schools do not do all of this. There are some similarities and differences in the testing consistency, but this leads to poor data. There are teachers and schools who cheat and there are teachers and schools who do not. If data is going to be good and accurate all Knox County Schools/state schools should be consistent in this method. Also, the curriculum is written to match what is on the state mandated test. We do not have a choice of teaching that or not. We never get our scores back in enough time to reflect on our teaching practices and more than half of us do not understand the language used and what the scores mean on the TCAPS. It is ridiculous.*

(Respondent 815) *Although the state mandated tests do provide valuable information on individual students, how this information is used is problematic. Are we using them to teach to the test? I believe this is true. We should be using them to create differentiated instruction so that every child feels successful as he/she reaches full potential. It has been my experience that children learn better when they work together and the teacher facilitates their learning. Anyone who thinks drill and skill teaching to the test works needs to spend some time in a classroom where this occurs. The blank looks on the children's and the lack of motivation to complete assigned tasks tell it all.*

Sociocultural Influences and Intervening Variables

Teachers responding to the open-ended question of effects relating to teacher work and identity named numerous sociocultural influences and intervening variables

occurring within their school settings. Teachers, recognizing that they are not responsible, nor could they be, for all aspects which influence student test scores teachers gave detailed accounts where intervening variables were simply out of the control or influence of the teacher. Teachers reported a loss of professional control and a sense of feeling powerless to make a difference when a highly influential yet unstable single test score is attributed to their work as a teacher. Factors of socio-economic concerns test score contamination, lack of parent support and inadequate resources are consistent with the comments of teachers who comprise this sample.

(Respondent 231) *I have collected considerable anecdotal evidence from students about how the teachers at their elementary schools assist them when they have a problem on the tests, including pointing out two answers to choose from out of the four possible choices, partially working problems on the board, reading questions, and so on. We receive all new students at my school in grade 6, and half of the students are new to us in grade 7. Interestingly, scores on the state-mandated tests are poor in grade 6 even among students who came to us with an advanced rating the year before, and the same is true in grade 7. And yet, in grade 8, scores suddenly increase dramatically, indicating that we have indeed been teaching something., I feel that the high stakes of the mandated tests have caused many educators, under intense pressure from the administrators and central office, to lower their personal morals in favor of keeping their jobs., A year ago I took the findings of my students responses to the writing prompt "What have your teachers done to make the TCAP testing easier for you to handle?" to the administration, and while some alarm was noted among the mid-level administrators, when they carried these concerns and student comments to the highest levels, they were told that it WAS NOT A PROBLEM and essentially to back-off. Needless to say, I have kept these comments against the day that my teaching career becomes threatened because students failed to show enough progress under my tutelage.*

(Respondent 276) *6 of the 8 years that I have taught were in 7th grade science. During those years, I discovered that parents and students should be held accountable as much as teachers are for the education of their*

child. However, all the responsibility falls on the teacher. That's why I now teach PE.

(Respondent 550) *I see so many kids who can't concentrate for the amount of time on tests. 60 and 65 minutes for reading and language arts tests are a long time for kids who do not like or have difficulty reading. In Knox County the reading and language are now separated areas, and the test are together. What is a problem to me is when the two areas were separate on the test, Knox County had the two together under one supervisor. Some schools teach reading and language arts together, but some teach them separately. Also there seem to be over lapping problems with test questions. , It seems to me there is one thing we are teaching for these days, and that is the test. If I have a good group of students who care about the test, I will do better. If I have kids and parents who do not care if they do their work or not, then I do not look so good because they will learn less in class. , The more I teach I see parents as the ones who also influence the scores. If my parents want their children to do well, they are active in their learning by checking their work, working well with teachers, and having positive attitudes about education and its importance.*

There seems to be less time for teachers to prepare and assess learning. How can we do our best if we have to do all of the work after school hours? I believe happy teachers make happy, teachable students. Praise is necessary for happy teachers so therefore we have happy students who learn. I think it is often forgotten that the teachers are the ones who work with the kids, and if we were respected more, there would be more learning.

Rewards and Sanctions

Teachers reported varying effects in relation to rewards and sanctions connected to their work and teacher identity. Overall, rewards and sanctions attributed to high-stakes testing were recognized by teachers as having negative and detrimental influences on their work. Specifically, teachers described a competitive and stressful work environment where effects of high-stakes tests impacted “who they were as a teacher”.

(Respondent 162) *I feel it is creating a competitive nature among teachers especially when tied to teacher and school incentives (i.e. like the TAPS program in Knox County). I think it has taken away teacher autonomy and reinforced scripted reading programs and direct instruction. Education is no longer well rounded but is about "teaching to the test." Sad--that's why I'm leaving education.*

(Respondent 163) *I have been threatened with an administrative transfer by my principal because of my value added test scores... scores from TWO years ago since we have never received scores from ONE year until the END of the NEXT teaching year. That was bad enough....what is worse is that our children are becoming extremely good at recalling information due to reteaching and review on all the standards throughout the year....but they cannot THINK their way out of a paper bag! Our first priority in education should be to teach children to LEARN. It doesn't matter if they get all the FACTS in one year. If they are DOING, they are learning...and more important. THINKING. Those are the skills they will need in real life...and NOT if they can remember the different types of symbiotic relationships.....unless of course they plan on being a contestant on "Are You Smarter Than A Fifth Grader?"*

(Respondent 628) *I find the test very stressful because most of my students freely admit that they do not try on the test (many don't even read the test selections) because they know the scores only affect the teachers and not them, yet my principal uses the test scores as a measure of our teaching ability.*

The inclusion of the narrative teacher voice stands as a corroborative, perspective view within a social and culturally influenced high-stakes testing environment. Narrative data contribute a rich and insightful teacher voice to the empirical data findings of this study. These narrative data utilized as a corroborative measure serve to minimize the potential of researcher and situational subjectivity bias in data reduction (Engle, 1984). Teachers responding (N=112) to the second research question present a remarkably similar respondent voice. Analyses of these narrative data sets through manual coding procedures and confirmatory computer assisted technologies (Text Analysis SPSS for

Survey Research software) has allowed the research to confidently incorporate the teacher voice as being representative of the range of teacher perception and experience in relation to the high-stakes testing environment.

Proponent and opponent view

Data were further categorized to determine a proponent or opponent respondent voice in terms of the effects of high-stakes testing on teachers' work and identity. Of the 112 respondents, only 2.7 % of the teachers included comments which may be attributed to a positive or supportive response in relation to the effects of high-stakes testing.

Overwhelmingly, 97.3 % of the respondents reported experiences which were highly characteristic of an opponent or detrimental view to their work and teacher identity. For example, teacher comments detailed negative teacher morale, teacher retention issues, high levels of stress, resentment, worry, competitive climate and loss of professionalism.

While these respondent teacher comments did not deny the important instructional component which standards and sound accountability policies provide to their work as a teacher, they did articulate effects of high-stakes testing which impacted and impeded their work and teacher identity. However, a small group of teachers (2.7 %) described positive experiences attributed to high-stakes tests such as greater focus and attention to: prescribed curriculum, accountability and schools with lower socio-economic communities. Although these teachers stated they had never seen test results, they did recognize the potential for test data to inform their instruction. Moreover, teachers responding to this research question reported significant negative and damaging aspects attributed the high-stakes testing environment.

Connectedness of the effects of high-stakes testing

The results of research question two are visually represented in the following model (Figure 4:01) depicting the numerical frequencies and characteristic similarities in how teachers responded. This visual model visually presents the frequency of the traits and illuminates the relational qualities of the themes present in the data.

Teachers responding to the optional open-ended research question were remarkably representative of a common teacher voice. This model portrays the connectedness of teachers' narrative responses. For example, while teachers may have begun their response specifically referring to one salient characteristic of their teacher experience they tended to narratively corroborate and name several additional characteristics to describe the highly relational nature of the effects of high-stakes testing impacting their work as a teacher. Results indicate that 41, or 36.7 % of these teachers responding, described effects of test preparation and curriculum narrowing on their work and identity as a teacher. Comments associated to teachers describing their feelings and roles within the high-stakes environment were 39, or 34.4 % of these respondent data. Teachers' responses concerned with test use and valuing of the test were 46, or 42.2 %.

Sociocultural influences and intervening variables which impacted teachers' work and identity were associated with 38 teacher responses, or 33.9 %. While teachers responded in detail and at length about the effects of high-stakes testing on their work and teacher identity, results showed that only 9, or .08 % of teachers responding, chose to discuss ways that rewards and sanctions have affected their work and teacher identity. Illustrating the significant themes within these narrative data responses this visual figure stands to

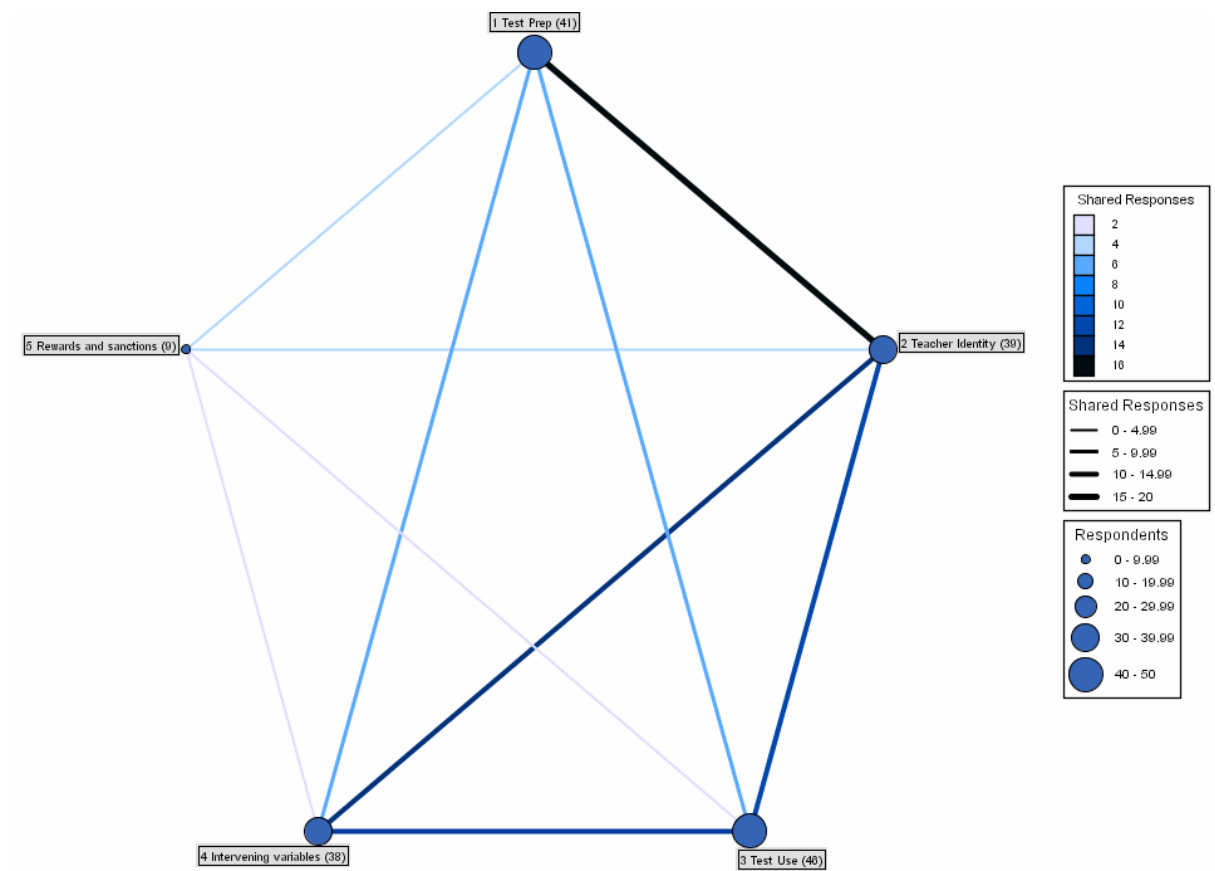


Figure 4:01: Connectedness of the effects of high-stakes testing

further depict the complexity of the effects of high-stakes testing on teachers' work and identity.

Results of Analysis

The results of this quantitative survey research on the effects of high-stakes testing on teacher pedagogy, practice and identity indicate a complex and paradoxical educational environment. This study addressed the complexity of the research questions where teacher perceptions of high-stakes testing were analyzed utilizing descriptive, correlational and corroborative data analysis techniques to answer the first research question. Descriptive and inferential statistics were employed in the analysis of data. Frequencies were computed for each survey question. Multiple correlations were computed and tested for significance. Narrative data were coded and analyzed for seminal and corroborative categories which both answered the second research question and further supported the results of the first section of the survey. Rigorous survey methods and research design further supported the reliability and validity of the findings. Finally, all results are interpreted within the limitations of self-report data.

Results Summary

Chapter four has presented the results of the analysis of an extensive quantitative survey study on the effects of high-stakes testing on teacher pedagogy, practice and identity. Analysis determined four factors emerged from the data: climate, pressure, alignment and value. Results of this analyses indicated that teachers in elementary

schools experience greater pressure related to the high-stakes test than teachers in middle or high school. Additionally, teachers in below average performing schools report great influence and impact of the high-stakes test on their pedagogy, practice and teacher identity. Finally, teachers representing over 25 % of the sample responded to the open-ended question corroborating the findings from the first section of the survey instrument that high-stakes testing does indeed affect teacher pedagogy, practice and identity. Chapter five will provide the discussion, conclusions and recommendations from this quantitative research.

CHAPTER 5

DISCUSSION

Chapter Introduction

High-stakes tests are the driving force behind the current federalized educational reforms (Abrams, 2004; Abrams & Madaus, 2003). The 1983 release of *A Nation at Risk: The Imperative for Educational Reform* (National Commission of Excellence in Education [NCEE], 1983) spawned the current high-stakes testing movement. This report called for a new educational purpose of rigor and academic success for all children. Today, the outcomes of this early report are played out in schools across America and scripted by the standards, assessment and accountability initiatives of the far-reaching legislation of the NCLB Act. To account for progress and determine academic success, the NCLB initiatives have required states to test students by means of tests which hold high stakes for all. With the advent of the NCLB Act, America's students are tested at a rate which has surpassed that of any other industrialized nation (Morrow, 2001). Also, the importance placed on high-stakes tests has increased exponentially since the minimum competency movement of the 1970s (Allington, 2002). Finally, with lead politicians repeating such aphorisms as, "What is tested gets taught", the voice of teachers becomes the voice of balance, expertise and responsibility to the purpose of schooling.

High-stakes tests critically impacting programs, pedagogy, practice, curriculum, individual student achievement, and teacher identity have proven to be a complex burden for teachers in terms of determining what constitutes quality education (Noddings, 2007;

Valli & Buese, 2007; Valli et al., 2008). While high-stakes tests are viewed as efficient producers of test scores, these scores are metaphorically held as reliable proof and valid products of teaching and learning (Corbett & Wilson, 1991; Cronin et al., 2007; McNeil et al., 2008). However, there is growing evidence which shows that high-stakes testing has resulted in educational effects which have not improved overall student achievement or met the primary outcomes of the NCLB mandates (Allington, 2003; Amrein & Berliner, 2002b). Both proponent and opponent views of high-stakes testing have found support in the literature (Au, 2007; Cimbricz, 2002). Findings from this study reveal that the consequential effects of high-stakes testing have had a powerful and profound effect over teachers' pedagogy, practice and identity.

Review of Research Rigor

Throughout this quantitative survey research, the researcher sought to provide sound and rigorous data analysis to support the conclusions and recommendations based on significant findings. Standards of rigor guided the establishing of internal and external validity, allowing the results of this study to be generalized with caution to a greater population. Internal validity issues were addressed utilizing representative sampling techniques, thereby ensuring the respondents were typical of the theoretical population as well as to within group and to the possible sample. The highly contemporary and relevant nature of the research question ensured that this quantitative research was not an individualized or special case, thereby ensuring greater external generalizability.

In summary, this quantitative survey research sought to maintain standards of research rigor and quality. Through the use of a quantitative survey methodology the researcher surveyed a representative sample of K-12 teachers in Eastern Tennessee to ask two significant research questions. Utilizing general principles for high quality, scientific research in education (Howe & Eisenhart, 1990), the researcher employed appropriate methodology to provide a logical and relevant chain of reasoning. The intended goals of this research were twofold. First, the research compared and examined the relationship of the consequential effects of high-stakes testing on teacher pedagogy, practice and identity. Second, critical to developing an understanding of teachers' perceptions and experiences within the current high-stakes educational setting, the researcher sought to document the "voice" of the teachers. Additionally, this study aims to provide a quality research contribution to policy formation and planning of future educational reform initiatives.

Discussion

This quantitative survey research was informed by sociocultural theories which provided relevant underpinnings and connections to the analysis of data, linking theory to practice. The review of the literature situated this current research within a broad disciplinary conversation regarding the effects of high-stakes testing. Additionally, the review of literature illustrated the importance and need for teacher voice to be documented and included in the research. Further, the utilization of quantitative survey methodology served as a 'best fit' to answer the research questions. Survey methodology

provided an anonymous response format for teachers who are caught within a highly politicized educational environment. This quantitative study examined teacher self reports of the effects of high-stakes testing asking two central research questions:

6. What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?
7. What are the consequential effects of high-stakes testing in relation to teachers' work and identity?

Teachers responding to survey question number one reported their perceptions of the effects of high-stakes testing in relation to their pedagogy and practice. Teacher responses were factored into four salient themes:

1. school climate
2. pressure
3. alignment
4. value of test

Additionally, teachers responding to the optional open-ended question number two reported their perceptions of high-stakes testing in relation to their work and identity as a teacher. Teacher responses to question two were categorized into five broad themes:

1. test preparation
2. teacher identity
3. test use and value
4. sociocultural and intervening variables
5. rewards and sanctions

This chapter will discuss the findings of these results. Additionally, conclusions based on the research findings will be presented. Finally, recommendations for future research and implications for teachers' work and policy formation will be explored.

Research Questions

The first research question was answered utilizing frequencies, correlation analyses and narrative corroborative data. The focus of this research question was to document the voices of teachers asking:

- *What are the consequential effects of high-stakes testing on teachers' pedagogy and practice?*

In responding to a series of Likert-type questions, teachers recognized the importance of meaningful and appropriate standards and accountability measures in relation to their pedagogy and practice. This characteristic of support for both standards and accountability measures stands as a salient feature of teacher responses to this question. Regardless of stance, the collective tone of teacher responses was one of experienced and committed professionalism. Teachers voiced their knowledge of the integral nature and importance of standards and accountability in relation to their work as teachers. However, while teachers expressed a support for standards and accountability in theory, many teachers reported that in practice, they held a significantly contrasted view of test scores than the federally stated view. Teachers stated they did not share the federal view which regards the ability of a single test score, acting as an effective lever of change, to guide or improve the quality of education. This contrasted view was a

contentious feature of the majority of teacher responses where high-stakes tests have become an increasingly controlling and constraining factor in their daily work (Valli & Buese, 2007). Overall, teachers responding to this research question answered in similar ways which were reflected in the review of literature (Barksdale-Ladd & Thomas, 2000; Hoffman et al., 2001; Pedulla et al., 2003).

Teachers stated that the heavy reliance on a single test score was, in their experience, a highly fallible and unstable measure of both teaching and learning (Haney, 2000). Teachers described test results as being highly sensitive to school population differences such as: socioeconomic differences, home resources and parental involvement. Teachers reported that these fundamental differences were the common complexities and inequities of their school settings and thus, the single, generic test was not an accurate measure of what students had learned. Moreover, teachers viewed test scores as a greater reflection of individual student background and the experiential characteristics students brought to the test experience rather than an accurate indicator of their teaching.

Teachers reported there are many intervening variables which may affect the outcome of high-stakes tests making the current practice of holding a teacher accountable to this single measure highly impractical. However, while teachers reported their caution in attributing great significance to a single test score, they stated that it is highly practical that they collect evidence of student learning throughout the year, resulting in a representative body of assessment evidence to both inform instruction and report on

student learning. (Valenzuela, 2004). Despite this practical awareness, teachers caught in this current accountability scheme see that:

(Respondent 367) *The only people held accountable are the teachers.*

While teachers recognized that tests, used as critical measures with high-stakes attached, have created highly test responsive school settings, they also recognized that high-stakes tests are an extremely limited and narrow measure of teaching and student knowledge. Worrisome to these teachers was the fact that – it all comes down to one test.

(Respondent 403) *Student success is based on a week of nonstop testing rather than the body of their learning.*

Again, even though these teachers expressed seemingly negative comments regarding high-stakes testing, it is essential to discuss an important characteristic of the complete data set of teacher responses. Throughout the data set teachers were open and candid, describing their frustration of the high-stakes testing environment but, overall, they expressed a commonly held support for both standards and accountability. Often, what teachers asked for was greater clarity and purpose in the standards and accountability mandates. More often, teachers asked to be included in the conversation. Perhaps, teachers, administrators and policymakers are in greater agreement in regards to the structures of standards, assessment and accountability than has been previously recognized on these issues. There is agreement among teachers that while accountability in the form of the high-stakes testing mandates have gone awry, schools without some kind of measure in place to assess teaching and learning would be schools without a place

for common dialogue, purpose and success (AERA, 2000). Teachers reported that the high-stakes test:

(Respondent 272) *Provides valuable information, but it cannot be the sole measuring stick. Sometimes it is an accurate reflection, but for some students it is simply not.*

(Respondent 383) *I believe that too much emphasis is placed on testing. I do believe that teachers should be held accountable, but that a single test score should NOT be the only indicator.*

Here, teachers are stating that the metaphorical “one size fits all” test is just not working. The current practices resulting under the accountability policies are seemingly counterproductive to improving academic success opportunities for all students. The study’s findings indicate a teaching population caught within a deteriorating educational climate. The dilemma here appears to be that the systems in place for accountability are out of sync with the needs of teachers and students. The problem requires a reconceptualization of who is accountable and how to include and establish a teacher voice in the move towards a practical and purposeful plan.

Teacher responses to items assessing this research question represent a piece of the conversation which needs to be heard. However, previous research has found that voices of teachers have often been “dismissed by testing advocates who argue that teachers oppose high-stakes tests simply because they do not want to be held accountable; teachers are biased so their concerns about high stakes tests should not be warranted” (Wright, 2002, p. 5). The findings from this study stand in contrast to the perspective view of teachers who do not value accountability and whose biases cloud

their pedagogy and their practice. These teachers, in this study, are clearly in support of standards and accountability, and at the same time offer a knowledgeable and highly experienced voice to working within the complexities of the high-stakes testing environment. This self report data is representative of the day-to-day experiences and concerns that teachers have with such a generic “one size fits all” test.

Teachers reported that the current high-stakes accountability measures have accomplished little to lessen the achievement gaps and address individual needs of students. Teachers responding to the open-ended question described a teaching environment which was guided by the test preparation and not by the instructional needs of their students.

(Respondent 424) *If it is not an SPI ... DO NOT teach it!*

Within this educational context, alongside of the current, complex nature of high-stakes testing within the local school settings, it will require a collective understanding and commitment of all involved to foster and develop pedagogy and practice aligned to commonly held, fair and quality academic standards at the local and site level. As a result of the push to control teaching and learning (Britzman, 2003) from a governance and regulatory perspective utilizing high-stakes tests - teachers reported they are not supported or enabled to do the responsive work of teaching (Johnson & Johnson, 2006).

(Respondent 307) *I feel like I am constantly throwing information at the students.*

Teachers responding to this study expressed a desire to participate in opportunities for professional dialogue and conversations surrounding the effects of high-

stakes testing. Teachers expressed their dismay of being excluded from the conversations which directly impacted their daily pedagogy and practice. Here, opportunities for teachers and administrators to engage in these important conversations may begin to alleviate the tensions expressed by teachers, especially in those schools designated as low performing schools.

The results of this study showed that demographic composition of schools played a significant role in determining how teachers experience the high-stakes testing mandates. One of the strongest factors determined from teacher responses was attributed to how teachers valued the high-stakes test. The teachers working in schools designated as below average performing schools reported that they held little value of a “one size fits all” test which has limited utility for the demographic, social and cultural experience of the children they teach. One may conclude from these findings that teachers believe that the instructional needs of students in schools designated as below average performing are vastly different from the instructional needs of students at average or above average performing schools; thus a generic, “one size fits all” test has decidedly limited utility and value for these populations.

Further impacting teachers’ work was the exponential effect of students high-stakes test scores on their work; teachers reported that as school test scores decrease the impact and pressure on teachers increases significantly (see Table 4:20). In exasperation, one teacher stated:

(Respondent 126) *Inner city kids will never test as well as (deleted reference to specific suburban community) kids EVER!! They lack prior experiences.*

Teachers who taught in suburban or rural schools reported a greater impact of the high-stakes test than their colleagues teaching in the urban school settings (see Table 4:18). These differences may be attributed to factors such as: a more diverse student population present in suburban and rural schools than present in the urban school setting and possibly less funding and resource allocation available in the suburban and rural setting schools in relation to urban schools. Several teachers reported that they did not find the same value that policymakers attributed to the test and are implied by test scores stating that:

(Respondent 693) *There are so many variables that can affect a given student's performance on any one day or set of several consecutive days that using the data to indicate anything more than a range of performance at that time is a stretch to me.*

Teachers reported that because test results are given so much value within the academic setting, test scores heavily influence their day-to-day pedagogy and practice. However, 17.6% of teachers from this sample reported (see Table 4:07) that they never see the test results. Teachers acknowledging that high-stakes tests are the measures of the day to which they are being held accountable - ask to be given the tests and the results in order to do the work required within a well supported reform initiative. This practice, of not revisiting test results, stands in contrast to a practice which may support the intended outcomes of increased student learning and improved test scores, While it seems logical that teachers who are being held accountable for the outcomes of these high-stakes tests

should be given the test and the resulting data to inform their pedagogy and practice, teachers reported that:

(Respondent 92) *I've never seen a copy of a test or any questions from a test.*

(Respondent 471) *We never get our scores back in enough time to reflect on our teaching practices.*

These comments positioned within the context of the research question beg for school districts to recognize that the issues of standards and accountability are not separate from the work of teachers' daily pedagogy and practice. If change, improved student learning and higher test scores are the goals, then for administrators to spend time in the data alongside of their teachers may be an accessible characteristic in meeting these goals (Darling-Hammond, 2004).

The teachers responding to this survey research represented K-12 urban, suburban and rural school settings. Findings determined that elementary teachers reported greater impact of the effects of high-stakes testing on their pedagogy and practice in terms of both school climate and pressure related to the test (see Table 4:17). This makes sense, as elementary teachers are responsible for all curricular subject areas including those designated as high-stakes tested subjects. Whereas, middle and high school teachers are not responsible for subject areas other than those specific subjects they have been assigned to teach. Elementary teachers report they are under immense pressure to prepare their students for all aspects of the high-stakes test. It is unlikely that middle school teachers responding to this survey research all taught curricular areas which were subject

to high-stakes testing. For teachers working within these kinds of work settings, where teachers are either in or out of the testing target, high-stakes accountability can prove to be a divisive force within a teaching community. The data show that middle school teachers responding to this survey are acutely aware of the specific high pressure curricular subjects and grades in relation to their pedagogy and practice.

(Respondent 574) *I teach fine arts, which is not state mandate tested.*

(Respondent 668) *The emphasis on high-stakes testing at the elementary level takes much of the joy and creativity out of the teaching profession and causes unwarranted stress for both students and educators.*

Early research (Jones et al., 1999) describes similar findings where elementary teachers have had to choose to de-emphasize non-tested curricular areas in order to prepare students for the test. For example, teachers responding to this survey consistently maintained that high-stakes tests limited and constrained their ability to teach to the individual needs of their students. More than 97% of these teachers reported (see Table 4:08) spending valuable instructional time preparing students for the test. More specifically, 90.7% of teachers reported (see Table 4:21) using instructional time to test and practice test-taking skills. Importantly, these teachers recognized that while they had to deal with constrained curriculum and the threat of sanctions, their students deserved more than the scripted test preparation which formed the curriculum of the day.

(Respondent 294) *The pressure teachers feel because of the high-stakes testing contributes to their feelings that they have no control in what they teach or how.*

These current findings are evidenced in the early literature by Allington and McGill-Franzen (1992); these researchers foreshadowed many of the factors teachers reported in relation to the effects of high-stakes tests as they impact the educational climate and contribute to the pressures of teachers' work. Teachers in this study described their teaching day as being highly constrained by activities defined as preparation for the test, with 65.9% of teachers reporting that they engaged in test preparation throughout the school year (see Table 4:09).

Findings further indicated that elementary teachers are significantly more likely to provide released test items to their students for test preparation (see Table 4:22). Data showed that the pressure teachers feel for their students to perform well on the test has created a teaching pedagogy that is bound and constrained by the content of the high-stakes test. Mirroring the findings of previous research (Amrein & Berliner, 2002b; Au, 2007) the teachers responding to this study recognize the constrained nature of the curriculum they are mandated to teach and worry if students will actually attain knowledge or value any concepts or ideas outside of what is tested.

(Respondent 259) *State testing has dramatically “dumbed” the English curriculum I taught 25 years ago.*

(Respondent 294) *Many (teachers) have given up the very strategies that promote critical thinking skills to teach to the test – skills they know will be on the test.*

Teachers reporting that a majority of their daily instructional time was devoted to test preparation describe school communities which are akin to test prep centers (see Table 4:10). As these teachers report:

(Respondent 534) *It seems we have created a generation of test-takers not necessarily critically thinking learners.*

(Respondent 125) *Too much emphasis on the outcome of standardized testing. We are teaching humans, not widgets.*

Teachers, in this study, described a loss of purpose when they spoke in terms of the prevailing belief that one can look to increases in student test scores as the indicator of effective teaching and improved student learning. Previous research has documented similar findings where teachers made pedagogical and practice decisions that resulted in teaching and learning environments which do not support the needs of the learners (Amrein & Berliner, 2002a). Teachers responding to this research described testing scenarios where the press for increased test scores overcame teachers' ethical and professional judgment.

(Respondent 521) *My teaching style has certainly changed and would make my college professors cringe! My student's value-added scores are among the best in the system because I know how to teach to the test!*

Findings show that as pressures and expectations of increased high-stakes test scores go up teachers who are impacted the most have turned to tactics and strategies which are contrary to their teacher knowledge. The results determined that teachers working in rural (see Table 4:23) and below average performing (see Table 4:24) schools provided more rewards for test completion than their colleagues working in suburban or

urban school settings and average or above average performing schools. These actions may be a greater consequence of the demographic background of the students in rural and below average performing schools, as these groups may represent a greater population of disadvantaged social and cultural backgrounds which tend to not wholly support or value tests. This assertion has been supported in the work of Jones and Egley (2004), stating findings that academic achievement of students is highly correlated to socioeconomic status. The data shows that teachers in these schools sought to use whatever means they deemed productive in order to raise test scores.

Concerns raised by teachers have been well evidenced in the literature regarding test score contamination (McGill-Franzen & Allington, 2006) and artificial inflation of scores (Allington, 2000). Teachers reported that students are increasingly coached and groomed to become effective test takers, spending a considerable amount of the instructional time completing practice activities aligned to the test format. However, a salient factor of test scores, and all numbers for that matter, is that test scores can increase “with or without real improvement in the broader achievement constructs that tests and assessments are intended to measure” (Linn, 2000, p. 4).

In contrast to a study conducted by Kennedy (2005) which reported that only 10% of teachers based their daily teaching pedagogy and practice to the test, teachers responding to this research have described a highly constrained pedagogy and practice which works inside of a highly structured accountability system (see Table 4:10). This “look-alike” teaching practice asks us to consider how much is too much in terms of test

preparation for a “one size fits all” test. In other words, while test scores may in fact increase, teachers report that this is more likely a result of the “look alike” test preparation and curricular alignment to test content rather than the outcome of effective teaching or the result of a rich curriculum and increased student learning. These higher test scores, although congruent with the goals of testing reform, run contrary to teachers’ professional knowledge and best practice (Valli & Buese, 2007). Teachers perceived that they were required to align their pedagogy and practice to the test (See Table 4:21).

Teacher responses to this research describe an educational environment where teachers are simply getting *the test* back in various iterations from their students. These findings are reflective of previous research describing the highly test responsive and homogenized curricular focus of teacher pedagogy and practice (Barksdale-Ladd & Thomas, 2000).

Teachers describe their teaching pedagogy and practice as being diminished and constrained by the high-stakes testing environment (McNeil et al., 2008). As a result, teachers view their role as purveyors of test preparation within highly test responsive school settings. Finally, teachers responding to this research question have indicated numerous effects of high-stakes testing that both control and constrain day-to-day pedagogy and practice.

The second research question was answered utilizing corroborative data analysis.

This research question asked:

- *What are the consequential effects of high-stakes testing in relation to teachers’ work and identity?*

Teachers answering back to this research question did so in highly similar ways.

Overall, the responses from this representative sample of Eastern Tennessee K-12 teachers were realistic and to a greater extent resilient within a complex and pressure filled teaching environment. Teacher responses were shaped by their lived experiences and included a sufficient sample size (N = 112) to determine confident and representative findings.

Caught within the press of high-stakes testing, teachers in this study reported a loss of “best practice” and “teacher knowledge.” Like Dewey (1916/1997), these teachers explain that what they know to be and what they stand for should not be outside of their day-to-day teaching practices. Teachers described their current instructional practices as being strategically directed away from a constructivist pedagogy and practice. Here, Britzman’s (2003) caution, that using mandates to “control learning, however, effects more than just the ... teacher’s practices. It also constructs views about knowledge and the knower” (Britzman, 2003, p. 225). One teacher worries:

(Respondent 527) *...testing practices force teachers to spend the majority of the academic time teaching in ways that do not produce engaged, intellectually functioning citizens.*

Teachers in this study reported a heavy reliance on skill, drill and constrained pedagogy. While these actions run contrary to the goals of student centered learning, these teachers do what they need to do and what they believe they can do. These findings are supported in the research (Barksdale-Ladd & Thomas, 2000; Mathison & Freeman, 2006; Valli et al., 2008); for these teachers the absence of professional autonomy and choice has created a highly test focused teaching environment where teacher knowledge

is rarely considered. While these teachers responded to the open-ended survey question with highly descriptive experiences that articulated a constrained and test responsive teaching environment, these teachers also presented another layer of teacher identity which was rooted in deeply held professional teacher knowledge. Teachers in this study viewed their work and identity in comparatively immediate and fundamental ways, where teachers voiced their sense of agency and emancipatory actions in terms of getting on and making do (de Certeau, 1984).

(Respondent 529) *Until someone finds the courage to expose how harmful this type of evaluation can be, we will continue to endure the folly of number crunchers who call themselves educators.*

Responses from these teachers represent what teachers are most concerned about in relation to their work and identity. While testing mandates have dramatically changed the work of a teacher, these teachers describe a counter or alternative identity as they strive to ethically “do good work” (Gardner, Csikszentmihali, & Damon, 2001). Making do for these teachers is neither an action of being complicit nor compliant. Representing a remarkably unified voice, these teachers work towards a democratic presence at the classroom level within the high-stakes testing environment (Day, 2004). However, within these teacher responses are also the tired voices of teachers who have been sidelined too often and for too long. The pressures and tensions that teachers report in relation to the test are complex and far-reaching. As a result, teachers’ work and career satisfaction is reflected in comments which are representative of teacher morale.

(Respondent 190) *I teach what is required by the state. I feel like if I teach the material, the students will do well regardless of how I teach. Thus far this has been true.*

While this strong sense of pragmatism, which undergirds many of the responses, may be construed, by some, as negative diatribe, others will find it indicative of the realistic and resilient nature of these teachers.

(Respondent 115) *I would find it more useful to receive information on areas that I should have taught and how students did and what kind of information that they are not answering correctly on the test. This way I can focus more time on it for instruction or take enrichment courses in areas where I am weakest.*

(Respondent 126) *It can be very useful when presented in a positive way. When it is used as a threat for students or teacher, it makes people nervous or bitter. I believe if we really want schools and teachers to use the information to reflect on their practices, we must free them from the negative effects and empower them to improve.*

Even though these teachers described their role as purveyors of test preparation, they maintained an ethic which would make any group which must work as a team hopeful.

(Respondent 142) *I try to include multiple disciplines when teaching art to make sure that I am helping with test scores as much as possible.*

The teachers in this study are committed and professionally able to engage in the messy conversations and work ahead. Teacher responses described an internal struggle, predicated upon the expectations and mandates of the testing reform and their professional knowledge and work of teaching. Taking a sideways glance at what teachers perceive to be the mandated and sanctioned work environment of teachers, it is evident

that change which includes the voice of teachers is required. Perhaps the urgency of a required policy shift here becomes more apparent when these teacher voices are shared:

(Respondent 58) *High-stakes testing has negatively impacted teacher morale and retention in my district. I know of several other teachers my age (20's) who have left the profession because of the pressure to get students to perform on the test.*

Recognizing the impact of excluding teachers voice from the conversation, findings from the Hoffman et al. (2001) study suggested that the consequences of the high-stakes accountability environment in Texas has led to excellent teachers leaving the profession. The voices of the respondent teachers of this current study echo the findings from research, conducted by Johnson and Johnson (2006) that described similar effects of high-stakes testing on teacher identity and attrition. These researchers found that teachers working under such high pressure working conditions will choose to leave the profession and are often recruited by other employers. “That’s exactly what many have done and what many more will do” (p. 17) as the pressures of standards, accountability and high-stakes continues to impede the work of teachers - more and more teachers will simply choose to leave the profession.

(Respondent 434) *If I were just beginning my career in teaching, I don’t know that I would stay.*

Given these candid responses of teachers, it was somewhat surprising to note how often teachers in this study referred to student test scores as “my” scores. Clearly, these responses indicate that while teachers are contemplating other options, they continue to be committed to their work as a teacher. Such open and candid telling, positions the

effects of high-stakes testing at the center of teacher identity. In contrast to the findings of Elmore and Furman (2001) that stated they did not find evidence of teacher internal accountability in the public, charter and independent schools they studied, the teachers in this study reported on many levels ways in which they felt accountable to the high-stakes testing mandates. Findings from this study yielded a teacher identity which is strongly committed to the role of teacher but clearly confused and overwhelmed within the fallout from the federal, state and district accountability systems. Here, the complex nature of teachers' identity within the social and cultural practices of the everyday life of teachers' work has resulted in a confused, pressure filled power relationship of converging teacher roles (Valli & Buese, 2007) within the reform agenda.

(Respondent 311) *However, the test reports that good or bad – it is all because of me.*

Teachers spoke of their commitment to their students as a powerful and firmly constructed part of their teacher identity. In some instances, teachers knew that while test scores were both the “carrot and the stick” they consciously chose to enact a teacher identity which was moved by a significant individual and collective recognition of self – as teacher (Danielewicz, 2001). Quite simply, teachers knew that such complex and profound outcomes could not be one teacher's sole responsibility. In terms of teachers' work and identity, teachers responding to this research question describe a highly constrained work environment where the social and cultural characteristics of teacher identity are not included. Teachers responding to this question struggle daily with the tensions between their own professional beliefs and knowledge about effective teaching

and the high-stakes testing mandates. Teachers are letting us hear – loudly and clearly – while they cannot be held solely accountable for the academic success of all children, they are here to do the hard work.

(Respondent 401) *We need accountability, with BALANCE, in order to teach others things such as critical thinking, group dynamics, problem-solving, good character, service to others and creative thinking.*

(Respondent 491) *I am going to teach to the best of my ability regardless of the testing.*

And, finally:

(Respondent 511) *It is one of the most ridiculous developments during my career.*

Conclusions

This quantitative survey research has several strengths. The representative sample provides a common voice representing teachers in Eastern Tennessee. Two levels of data collection, Likert response and an open-ended question, yielded data with high internal validity in relation to the effects of high-stakes testing. Further to this, analysis was framed by Stone's (2002) epistemological stance of numbers as metaphors. The numerical data represented teachers' self reporting of their experiences and perceptions of the high-stakes testing environment. Typically, when presenting results numerically, there is an assumption of a common ground in understanding what those numerical classifications represent. This study recognized numbers or scores, if not described fully or grounded in meaning, as rendered metaphorically up for interpretation. Each number or score represents a highly qualitative and full experience within a range of measure.

The research reported here is complex and highly representative of a typical school setting. The findings of this research are less complex and call for reaction at the local and state level. The results indicate there is a significant relationship between high-stakes testing and teacher pedagogy, practice and identity. These findings have considerable implications for teachers at the local level, for schools' administrative staff looking to support teachers as they work within the high-stakes testing environment, and for policymakers who purport to represent the greater public and purpose of education. As a result, these pervasive patterns in the data are even more disconcerting when the effects of high-stakes testing are considered in relation to teachers' work and identity. Finally, in order for any recommendations to be implemented and integrated, they must be rooted at the practical knowledge level of teachers.

(Respondent 508) *A good teacher will teach the curriculum to the best of her ability, molding the lessons and strategies to the students' needs. A good teacher will have an excitement and enthusiasm about teaching and the subject matter that will be contagious to the students. This is done on a daily basis... the test is a by-product. Having said that, I cannot begin to explain to someone not in the classroom the apathy from parent and student alike regarding education as a whole. My heart is broken daily*

There are several generalizations which can be concluded from the findings of this study. First, the results indicate that teachers' perceptions of high-stakes testing do indeed shape their pedagogy, practice and identity. Earlier research conducted by Richardson identified the importance of teachers' beliefs and how these beliefs strongly influence their pedagogy and practice within high-stakes settings (1997). For example, teachers in this study report that they are not supported to engage in the effective or best

practice teaching style which they describe as the foundation of their teacher knowledge. Teachers reported settling for a narrow curriculum where a “one-size-fits all” perspective has become the common pedagogy and practice of everyday classroom work (see Table 4:21). However, at the same time, teachers reported a level of subversive teaching which they felt disrupted the prevailing assumptions and tried to do what they could to teach beyond the controlled curriculum (Britzman, 2003).

(Respondent 673) *Teachers at my school are fortunate that the fire marshall has required us to stop using door stops to prop our classroom doors open; it is MUCH easier for us to teach in ways that we know are best for kids when our doors are shut.*

Teachers, in this study, who responded with some degree of dissatisfaction to the status quo, will not be surprising to those who are close to the work of teachers. Teachers recognizing the state of educational pedagogy and practice view the current mandates in terms of another political metaphor to demonstrate that something is being done (Dorn, 1998; Jones et al., 2003).

(Respondent 104) *The “one size fits all” approach to evaluating students and teachers is just another political stunt enacted by politicians and bureaucrats who are grossly ignorant of all aspects of learning.*

Second, and more worrisome, are the teacher comments that describe a barren professional climate which makes little or no room for purposeful professional development. However, this perspective also emphasizes a highly hopeful aspect of the responses from this study which describe a teaching population with a new “R.” While, these teachers describe their work and identity as being characteristic of rebels, renegades and resisters – there is a definite tone of resilience among these teachers.

Certainly, even following a theoretical business model upon which to work with the messiness of change within this highly structured curricular reform, the abysmal job thus far could be greatly improved with a workforce such as these teachers. The challenges at the state and local levels will be large; a common purpose of those who do the work is required to develop and effect meaningful organizational and system change. This is the messiness of organizational reform where teachers and administrators will be required to work together with a common understanding and purpose to meet the proficiency expectations of the NCLB high-stakes testing mandates (Valli et al., 2008).

Teachers responding to this research recognize that it does take time to think about complex issues. This study has situated itself within the tensions and complexities of the high-stakes testing environment where the “effect of too many policy demands coming too fast often resulted in teacher discouragement, role ambiguity, and superficial responses to administrative goals” (Valli & Buese, 2007, p. 520). The press to assess, evaluate and account for student achievement and teacher effectiveness has resulted in a highly test responsive teaching and learning environment where those in control have thought too hard and acted too quickly when it comes to “failure” (Allington, 1994b). Teachers who responded to this survey know they simply cannot accomplish what all other previous reform efforts have failed to do - close the achievement gap for all children (Abrams et al., 2003). Rather than view this act of teaching as the work of many, it is stoically grounded in the assumption that increased student achievement is the sole work of the classroom teacher.

(Respondent 679) *The only way we are going to make a real difference ... is to INCLUDE PARENTS AND STUDENTS in the ACCOUNTABILITY process somehow.*

The findings from this research highlight teachers' support of standards and accountability. In contrast the findings confirm their frustration and dissatisfaction with the system. Teachers in this study expressed the overall negative effects of high-stakes testing on their work and identity as teachers. Not surprisingly, teachers reported a loss of purpose under the disproportionate amount of pressure resulting from the explicit requirements of the mandated programs. As standards and accountability measures have become more influenced by outside forces, high-stakes tests at the classroom level have become increasingly constraining to teacher's work and teacher identity.

Ultimately, the tone of teachers' responses presents a duality of teacher identity; for example, the teacher who is under pressure to produce higher test scores is complicit and compliant in following the instructional guidelines and use of mandated resource materials (McNeil et al., 2008). And then, the teacher who seemingly is both complicit and compliant but is the rebel, the resistor, the renegade filled with the resilience of knowing that the system operates outside of who they are and who they choose to be as teachers. For those engaged in the work of reform, these teachers enact what many researchers have recognized as a highly resilient teacher identity. These teachers working within the highly constrained and structured high-stakes testing environment recognize power, position their work and move within those agentive moments (Holland et al., 1998). For these teachers, even though they are experiencing an immense amount of

pressure related to high-stakes tests, they continue to strive to meet the standards and accountability.

(Respondent 534) *Instead of the attitude of “this is what you need to know for the test”, I like “This is what you are going to learn about today!” Unfortunately, because so much rides on the test performance, it is very hard to maintain that attitude.*

Several conclusions from the findings of this study seem apparent but, because of the nature of the pervasive mandates of the high-stakes testing environment, none seems more important than to ask – what is the purpose of schooling? Once again, early research seems to have foreshadowed where we have arrived; twenty years ago McDermott (1987) stated, “The fatal flaw in American schooling will not be found in the reasons for one group failing in school...the ascription of failure to one person or group rather than another says nothing about the learning potentials of the persons involved” (pp. 363-364). McDermott cautions the dominant culture, arguing that, “We must be wary of our powers of articulation and explanation when they can keep us systematically dumb about ourselves” (p. 362). McDermott describes a system where broken notions of wrong doing and narrowly controlled curriculum have prevailed over the work and the identity of teachers.

This current research sees hope that we can look to previous research and educational practice to recognize ourselves caught in the muddle of the current accountability scheme and system failure. As McDermott articulated so many years ago, we have no time left for new theories of school failure and new ways to blame or defend. I am convinced that these resilient teachers are there to acknowledge the challenges and

weigh the opportunities to create a community that accepts difference rather than one which transforms “small and generally uninteresting differences in test-defined learning into institutional facts with devastating consequences for” (p. 364) all.

Drawing upon the findings from this study, it is evident that teachers do not hold a definite or clear understanding of the purpose of testing. Findings showed that teachers were more inclined to be test overseers rather than teachers. While teachers articulated they knew they were required to prepare their students for the test, they were specifically aware of the amount of time called for to prepare students for the test, and they understood the rewards or sanctions which could befall them if they failed in these roles; they were overwhelmed and confused in their role as teacher. What these teachers were unable to articulate was a connected purpose of the test to their work as – teacher. Ultimately, teachers voiced that they recognized the test as a metaphor for something other than good teaching and learning and seemed to view it as separate from their teacher identity.

In response to these teacher perceptions of the effects of high-stakes testing, the theoretical and practical importance of this study is substantial. It is clear that while much of the NCLB intent calls for an equal opportunity and education for all children, the effects of these reform mandates have created a muddle of teachers’ work. Teachers in this study state that schools have become test-driven buildings, refer to their students as “bubble kids”, and express awareness that in these current practices the purpose of schooling is lost.

In 1994 Allington (1994b) took a historical look back at educational policies and practices; he recognized then that the turn in the road for education began when the federally funded initiatives for “increased literacy proficiency and educational attainment were seen as necessary for furthering the development of the economy and the citizenry” (p.98). For schools operating today, where test scores and cost effectiveness are talked about in the same sentence, it is not the least bit surprising to this researcher that teachers find themselves caught in the press of producing and reporting inflated test scores. As has been widely referenced in the literature (Haney, 2000; Nichols & Berliner, 2007; Valli et al., 2008), as the stakes attached to testing increases so do the effects of the high-stakes test.

As discussed, findings from this quantitative survey research found that elementary teachers were the most impacted by the effects of the high-stakes testing environment. For example, teachers reported that they tended to direct a significant amount of their instructional time to preparing students for the test, providing content and materials which were directly aligned to the test, and offering rewards for test completion and achievement. Elementary teachers teaching in below average performing schools and teachers working in rural schools all reported spending less curricular time on those subjects which were not tested, less time on enrichment opportunities, assemblies, field trips and student choice time (see Table 4:25). With such an over emphasis on test scores, teachers report having to make the hard choices to displace other curricular experiences in order to prepare for the test. These kinds of pressures have caused teachers to feel they

have little control, reduced professional autonomy and less power over the daily kinds of pedagogical and practical decisions that make up teachers' work. Again, teachers speak of the many intervening variables which can impact a student's test scores and which ultimately revert back to a judgment of the work of the teacher. These excerpts describe a diminished sense of self as teacher within a work setting which offers little support:

(Respondent 551) *I strongly disagree with high-stakes testing in relation to my work and identity as a teacher because I am being judged on the work of a nine year old child. If that child came to school in a bad mood because they slept on the floor, woke himself up for school, did not have breakfast, and worries how their parents are going to pay the electric bill, I strongly believe that will affect the way the student will perform on the test that day. It is unfair to teachers.*

Teachers responding to this research have articulated what previous research has documented - all children simply will not "fit" within the standard cost-effective curriculum model (Allington, 1994b). Education is a messy business. In some sense, NCLB has cleared the floor and focused teachers' attention to the opportunities for educational success for all children. A current and more in-depth look would notice the preoccupation of blaming and defending when test outcomes do not match the set proficiency standards. Dare we work together to foster an educational setting where strengths and weaknesses are valued within an educational community of difference?

The focus of this research on the consequential effects of high-stakes testing allowed for an important question in any successful reform effort to be asked – How do high-stakes tests affect the work of teachers? When asked to discuss the effects of high-stakes testing teachers chose to discuss many of the consequential effects which have

undermined their professionalism. A significant implication arising from these findings suggests that teachers will either leave the profession or choose not to join the work.

These findings, depicting a loss of professionalism, a sense of being under valued and a loss of professional identity have been previously reported in the work of Hoffman et al (2002), Day, Eliot and Kington (2005), Lasky (2005) and Jones et al. (1999).

Increasingly, these teachers report they are challenged beyond their professional and personal capacity. Similar findings were reported by Finnegan and Gross (2007) suggesting that an unintended consequences of the high-stakes mandates has been teachers loss of morale when teachers caught in the press to conform and comply with test expectations have been unable to meet the test score targets. As a result, the teachers in this study report a teaching environment which enacts rewards and sanctions rather than creating opportunities to develop teacher capacity, expertise and success. Does this suggest then that the already short supply of teachers in America will continue to be affected by high-stakes testing? Could policy makers and school district administrators learn something from the findings of this study?

In conclusion, further research which is inclusive of the voice of teachers needs to be undertaken. Previous research has reported similar findings (Pedulla et al., 2003) of the consequential effects of high-stakes testing which are consistent with the experiences of teachers in Eastern Tennessee. Key findings from this current study will contribute significantly to the understanding of the effects of high-stakes testing and provide future research studies an authentic and foundational teacher voice. Similar to the recent

findings of Loeb, Knapp and Elfers (2008) these teachers have responded to the survey instrument with a remarkably unified voice, stating that high-stakes tests has influenced the content of what they teach and impacted how they teach. The results suggest that teachers are aware of the system of power and how their layered and multiple identities are shaped and recognized within their work as a teacher (Freire, 1995). Teachers responding to this research stand at the edges, ready and capable to assert themselves, filled with the “disappointment at the absence of revolution” (Habermas, 1995, p. 116).

The results of this study are conflicted with the basic misconception that teachers lack the motivation to do the work required of the NCLB Act (Valli et al., 2008). The results show that even though these teachers struggle with the tensions of their everyday work environment, these are teachers who recognize power, agency and position as it relates to their enacted role as teacher (Foucault, 1972). These teachers, while frustrated and overwhelmed, ask to be heard, to be counted and most importantly to be valued as – a teacher.

Recommendations

Overall, this study finds its relevance in the respondent teacher knowledge which has been documented in relation to the effects of high-stakes testing on teacher pedagogy, practice, and identity. This study may, in fact, describe the single and most important long standing variable to the success of any educational reform which simply states: “any major change, may hinge more on teachers’ perceptions of the change than on its actual merits” (Salvaterra & Adams, 1995, p. 35). The importance of including teachers within

the conversations and implementation of the reform efforts is perhaps the central issue in developing appropriate pedagogies and practices which support appropriate standards of learning. Unless teachers are included wholeheartedly they are unlikely to fully understand and support reform efforts (Turner, 2001). Further to this perspective, Richardson's (1997) research echoes the importance of teacher belief and perception in the success and implementation of any educational reform initiative.

Recent research conducted by Darling-Hammond (2004) maintains that the "issue of standards and accountability cannot be separated from issues of teaching, assessment, school organization, professional development, and funding" (p. 1081). According to the teachers surveyed in this study, the effects of high-stakes testing have gone largely ignored in the press for greater achievement and fulfillment of sweeping reform mandates. The failure to recognize these effects is undermining the vital and important aspects of the NCLB reform initiatives. Those who are in the lead must "refocus education policies to place a greater emphasis on supporting and improving teaching and learning, rather than relying on a system of rewards and sanctions to spur change" (Abrams et al., 2003, p. 27). This recommendation is based on the findings of this study and highlights the importance of how teacher professional development can develop capacity and provide shared learning opportunities for teachers as they engage in the work of teaching within a reform agenda.

In a recent interview (Crow, 2008) Richard Elmore, speaking of effective teacher professional development, maintained that what occurs outside of the local school setting

has little immediacy and importance for the classroom teacher. Elmore further contends that “powerful professional development occurs in real time around real problems in real schools involving real people who actually have to make decisions about what to do on a day-to-day basis” (p. 43). Together, Elmore (2004) and Fullan (2007) call for educators to move away from the model of professional development and adopt a stance of professional learning. These scholars realize that the hard work involved in school reform must be inclusive of teachers who have a strong sense of expertise and thus will develop as highly influential forces in meeting the goals of effective reforms. Teachers learning in practice creates a shared learning experience which in turn supports the development of an instructional community where teachers come together to learn and work together (Wenger, 1998). This view departs from the current climate of schools where teachers are viewed as purveyors of test content material preparing students for a test that is not used to improve the pedagogy and practice of teachers working with children.

Within this complex and politically driven educational setting there seems to be an inverse relationship between “commitment to what should be changed” and the “knowledge about how to work through the process of change” (Fullan, 1991, p. 95) for those holding fast to the notion that the teacher is solely responsible for the effects related to high-stakes testing (Spring, 2004). In contrast to this view of teachers as purveyors of test prep, Elmore points out that “there are models of accountability in ... which schools are given feedback on their performance of their students, and they are given support and

challenges to improve” (Crow, 2008, p. 46) their pedagogy and their practice in relation to the needs of the students they teach.

Recognizing the essential role that teachers’ beliefs and attitudes play in the success of any change or reform and knowing how teachers might respond to and engage in the reform agenda at the local and state level is key (Richardson, 1997). The findings from this current study support teachers becoming “active agents in the testing debate” (Smith & Fey, 2000, p. 343) who understand and are responsive to the social and cultural intentions which undergird educational standards and accountability measures. An important corollary recognizes that teacher identity and identity formation is clearly a unique and complex outcome of teachers’ experiences and their actions within the complex high-stakes testing environment. Understanding that high-stakes test results are highly site sensitive and that “teachers do not experience and respond to such policies in predictable, mechanistic, unidimensional ways” (Sloan, 2006, p 145) further describes the highly idiosyncratic and non-homogeneous work of teachers.

The research reported here is complex and highly representative of a committed and experienced teacher voice. The findings of this research are less complex and beg to be unraveled. Future research questions need to be asked and answered. Questions arising from this data ask:

- How can teachers become involved in the educational conversation?
- What counts as a relevant measure of knowledge?

- What factors foster and support an effective teacher pedagogy, practice and identity?

The results of this study are shared in the context of the current social and cultural educational environment. Previous research supports the findings from this study that high-stakes testing does indeed affect teachers' work (Abrams et al., 2003; Barksdale-Ladd & Thomas, 2000; Jones et al., 2003). Teachers' work is the labor of our social and cultural teachings and learning. The data suggests that teacher identity is reflexive and resistant within this educational environment. Some may argue that teachers need to adjust to the rigors of the NCLB reform and simply get on board with the reform agenda (Carnoy & Loeb, 2002; Firestone et al., 1997). Others may argue that the cost to teacher pedagogy, practice and identity has been too great (McNeil et al., 2008; Solorzano, 2008). The findings from this study indicate that the current high-stakes testing policies have served to devalue professional knowledge, distort the act of teaching and undermine both the standards and accountability policies in relation to teachers' work and identity.

Regardless, of one's position, it is recognized that these high-stakes tests and the resulting test scores have impacted the work of teachers in profound ways. As Wright (2002) points out, "[T]eachers are stressed and overwhelmed by all of the curricular changes and pressure to teach to the test and raise test scores. ... They are insulted when monetary rewards are disbursed to schools And they are frustrated when they watch good teachers leave" (p. 12) the profession. While tests are intended to act as a motivator to ensure effective teaching, the high-stakes testing measures overall have had the opposite effect. Based on the results of this study teachers maintain that the authoritarian

nature of high-stakes tests have worked to deprofessionalize teaching, have impacted their work and identity as a teacher and have resulted in negative influences affecting pedagogy and practice.

Moreover, results of this study indicate a significant relationship between high-stakes testing and the test-driven decisions teachers make about what and how they should teach (Hoffman et al., 2001). Teachers reported that the overemphasis and harmful impact of high-stakes testing “is affecting instruction in negative ways” and “is leading both students and teachers to ‘drop-out’” (p. 490). Results show that the high-stakes testing focus has had significant impact in narrowing curriculum and pressing teachers to focus on test related subject areas and de-emphasize those curricular areas and subject content which are not tested. Teachers recognized the high-stakes tests as powerful gatekeepers of the knowledge and skills to be taught (Jones et al., 2003). Teachers responding to this research identify a barren test responsive curriculum which has tended to drop away curricular subject areas which are outside of the test mandates. Teachers argue that there should be less time, money and effort spent trying to find the “silver bullet” or the “best” teaching guide and simply put a greater belief and investment in the professional judgment and practice of teachers.

Furthermore, as this analysis has shown, in order for teachers to do the hard work of teaching it is essential that educational reform moves solidly away from the deficit model it currently operates under. Recently, Lee (2008) noted that accountability must begin at the local level before external accountability can be both meaningful and

successful. Among the important findings from this study, is the recognition of teacher perceptions and actions as powerful and influential characteristics for the success of such a complex reform agenda. For example, developing and fostering local teacher commitment, capacity and expertise must be an important and integral part of a successful accountability system and plan. Effective reform will work slowly from goals and identified needs situated at the local level, recognizing the capacity and potential of the teachers in each and every individual school. As one clear voice articulated:

(Respondent 482) *Let us teach!*

This research has focused towards those paradoxical moments between mandates and practice where teachers' authentic voice is documented within an anonymous, internet survey methodology describing the lived experience of these respondents. Overall, these research findings may be reflected in Gramsci's (1971) notion of moving common sense to "the realm of Good sense" (Britzman, 2003, p. 69) and back again to the commonly held sense through the process of renovating the praxis between mandates and practice – assumptions and lived lives. The findings of this study contribute important dimensions of teacher voice describing the consequential effects of high-stakes testing. This study documents a teacher voice grounded in practical experience and knowledge that has been largely absent from the conversations regarding the profound effects of high-stakes tests on teacher pedagogy, practice and identity. These effects of high-stakes testing mandates for teachers and students ring sadly close to, "For whom the bell tolls" ... failure a priori.

And, one last question – do we want to continue as things are?

Of all the chasms that separate one world from another, none is greater than the gap between the people who make policy and the people who suffer the consequences... it will not do to sacrifice children on the altar of accountability, to use them in a giant high-stakes experiment and ignore the real harm it does (Kohn, 2002, p. 1).

LIST OF REFERENCES

- Abrams, L. M. (2004). *Teachers' views on high-stakes testing: Implications for the classroom*. Tempe, AZ: Education Policy Studies Laboratory.
- Abrams, L. M., & Madaus, G. F. (2003). The lessons of high-stakes testing. *Educational Leadership*, 61(3), 31-35.
- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory in Practice*, 42(1), 18-29.
- AERA. (2000). *AERA position on high stakes testing in pre-k - 12 education*. Washington, D.C.: American Educational Research Association.
- Allington, R. L. (1994a). The schools we have. The schools we need. *The Reading Teacher*, 48, 14-28.
- Allington, R. L. (1994b). What's special about special programs for children who find learning to read difficult? *Journal of Reading Behavior*, 26(1), 95-115.
- Allington, R. L. (2000). Letters: On special education accommodations. *Education Week*, 19, 48.
- Allington, R. L. (2002). *Big brother and the national reading curriculum: How ideology trumped evidence*. Portsmouth, NH: Heinemann.
- Allington, R. L. (2003). High-stakes testing does not improve the teaching or learning of reading. *The New England Reading Association*, 39(1), 31-36.
- Allington, R. L., & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York. *Educational Policy*, 6(4), 397-414.
- Alsup, J. (2006). *Teacher identity discourses: Negotiating personal and professional spaces*. Mahwah, NJ: Erlbaum Associates, Inc.
- Amrein, A. L., & Berliner, D. C. (2002a). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18) [On-line], <http://epaa.asu.edu/epaa/v10n18>.
- Amrein, A. L., & Berliner, D. C. (2002b). The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes

- results and ACT, SAT, and AP test results in states with high school graduation exams. [On-line], <http://edpolicylab.org>.
- Assaf, L. C. (2005). Exploring identities in reading specialization program. *Journal of Literacy Research*, 37, 201-236.
- Assaf, L. C. (2008). Professional identity of a reading teacher: Responding to high-stakes testing pressures. *Teachers and Teaching*, 14(3), 239-252.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Bakhtin, M. (1981). *The dialogic imagination* (C. Emerson & M. Holquist, Trans.). Austin: University of Texas Press.
- Barksdale-Ladd, M. A., & Thomas, K., F. (2000). What's at stakes in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education*, 51(5), 384-397.
- Baumann, J. F., & Bason, J. J. (2004). Survey Research. In N. K. Duke & M. H. Mallette (Eds.), *Literacy research methodologies* (pp. 287-307). New York: The Guilford Press.
- Baumann, J. F., Hoffman, J. V., Duffy-Hester, A. M., & Ro, J. M. (2000). The first R yesterday and today: U. S. elementary reading instruction practices reported by teachers and administrators. *Reading Research Quarterly*, 30, 934-956.
- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.
- Berry, B., Turchi, L., Johnson, D., Hare, D., Owens, D., & Clements, S. (2003). *The impact of high-stakes accountability on teacher's professional development: evidence from the south*. Chapel Hill, North Carolina: Southeastern Center for Teaching Quality, Inc.
- Bourdieu, P. (1977). *Outline of theory and practice* (R. Nice, Trans.). Cambridge, UK: Cambridge University.
- Bourdieu, P., & Passeron, J. C. (1996). *Reproduction*. London: Sage.

- Bracey, G. (2000). High stakes testing. *Center for Education Research, Analysis, and Innovation*(CERAI-00-32) [On-line], <http://epicpolicy.org/files/cerai-00-32.htm>.
- Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Educational Evaluation & Policy Analysis*, 12(1) [On-line], <http://epaa.asu.edu/epaa/v12n1>.
- Brewer, M., & Gardner, W. (1996). Who is the "we"? Levels of collective identity and self-representations. *Journal of Personality and Social Psychology*, 71, 83-93.
- Britzman, D. P. (2003). *Practice Makes Practice: A Critical Study of Learning to Teach*. Albany, NY: State University of New York Press.
- Callahan, R. (1962). *Education and the cult of efficiency*. Chicago: University of Chicago Press.
- Campbell, D. T. (1975). On conflicts between biological and social evolution and between psychology and moral tradition. *American Psychologist*, 30(12), 1103-1126.
- Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientists. In W. Trochim (Ed.), *Advances in quasi-experimental design and analysis: New directions for program evaluation* (pp. 31). San Francisco: Jossey-Bass.
- Carnoy, M. (2005). Have state accountability and high-stakes tests influenced student progression rates in high school? *Educational Measurement*, 24(4), 19-31.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis *Educational Evaluation & Policy Analysis*, 24(4), 305-331.
- Cawthorn, S. W. (2007). Hidden benefits and unintended consequences of No Child Left Behind policies for students who are deaf or hard of hearing. *American Educational Research Journal*, 44(3), 460-492.
- Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives*, 10(2).
- Clandinin, J. (1985). Personal practical knowledge: A study of teachers' classroom images. *Curriculum Inquiry*, 15 (Winter), 361-385.

- Clandinin, J., & Connelly, F. M. (1995). *Teachers' professional knowledge landscapes*. New York: Teachers College Press.
- Clarke, M., Shore, A., Rhoades, K., Abrams, L. M., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy, Boston College.
- Cochran-Smith, M. (2003). Learning and unlearning: The education of teacher educators. *Teaching and Teacher Education*, 19(1), 5-28.
- Cohn, M. M., & Kottkamp, R. B. (1993). *Teachers: The missing voice in education*. Albany: State University of New York Press.
- Conley, M. (2007). Reconsidering adolescent literacy. In M. Pressley, A. K. Billman, K. H. Perry, K. E. Refitt & J. M. Reynolds (Eds.), *Shaping literacy achievement* (pp. 77-97). New York: Guilford Press.
- Corbett, H. D., & Wilson, B. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464-494.
- Coyne, M. D., & Harn, B. A. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the School*, 43(1), 33-43.
- Crawford, J. & Impara, J.C. (2001). Critical issues: Current trends, and possible futures in quantitative methods. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 133-173). Washington, D. C.: AERA.
- Cresswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed method research*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, California: Sage Publications.

- Crocco, M. S., & Costigan, A. T. (2007). The narrowing of curriculum and pedagogy in the age of accountability. *Urban Education*, 42(6), 512-535.
- Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G. G. (2007). *The proficiency illusion*: Thomas B. Fordham Institute.
- Cross, C. T. (2004). *Political education: National policy comes of age*. New York: Teachers College.
- Crow, T. (2008). Practicing professionals. *National Staff Development Council*, 29(2), 42-47.
- Cuban, L. (2007). Hugging the middle: Teaching in an era of testing and accountability, 1980-2005. *Education Policy Analysis Archives*, 15(1) [On-line], <http://epaa.asu.edu/epaa/v15n1>.
- Danielewicz, J. (2001). *Teaching selves: Identity, pedagogy, and teacher education*. Albany, New York: State University of New York Press.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, 106(6), 1047-1085.
- Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many*. New York: Routledge Falmer.
- Davidson, A. L. (1996). *Making and molding identity in schools: Student narratives on race, gender and academic engagement*. New York: SUNY Press.
- Day, C. (2004). School reform and transitions in teacher professionalism and identity. *International Journal of Educational Research*, 37(8) [On-line], http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VDF-4BFXYJ8-1.
- Day, C., Elliot, B., & Kington, A. (2005). Reform, standards and teacher identity: Challenges of sustaining commitment. *Teaching and Teacher Education*, 21, 563-577.
- de Certeau, M. (1984). *The practice of everyday life* (S. Rendall, Trans.). Berkeley, California: University of California Press.

- Dewey, J. (1922). *Human nature and conduct: An introduction to social psychology*. New York: The Modern University.
- Dewey, J. (1938). *Experience and education*. New York: Touchstone.
- Dewey, J. (1997). *Democracy and education*. New York: The Free Press. (Original work Published 1916).
- Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method* (2nd ed.). New Jersey: John Wiley & Sons, Inc.
- Dorn, S. (1998). The political legacy of school accountability systems. *Educational Policy Analysis Archives*, 6(1) [On-line], <http://epaa.asu.edu/epaa/v6n1.html>.
- Downing, S. M., & Haladyna, T. M. (1996). A model for evaluating high-stakes testing programs: Why the fox should not guard the chicken coop. *Educational Measurement: Issues and Practice*, 5(1), 5-12.
- Dressman, M., & McCarthy, S. J. (2004). Towards a pragmatics of epistemology, methodology, and other people's theories in literacy research. In N. K. Duke & M. H. Mallette (Eds.), *Literacy research methodologies* (pp. 322- 346). New York: Guilford Press.
- Duffy, G. G. (2007). Thriving in a high-stakes testing environment. *Journal of Curriculum and Instruction*, 1(1), 7-13.
- Elmore, R. (2004). *School reform from the inside out*. Cambridge, MA: Harvard University Press.
- Engle, M. (1984). Factors affecting the reduction of narrative data: ERIC Document Reproduction.
- Farkas, S., Johnson, J., & Duffet, A. (2003). *Stand by me: What teachers really think about unions, merit pay, and other professional matters*. New York: Public Agenda.
- Fink, A. (1995). *The survey handbook*. Thousand Oaks, CA: Sage.
- Finn, C. E., & Petrilli, M. J. (2007). Forward. In *Proficiency illusion*. New York: Thomas B. Fordham Institute.

- Firestone, W. A., Goertz, M. E., & Natriello, G. E. (1997). *From cashbox to classroom: The struggle for fiscal reform and educational change in New Jersey*. New York: Teachers College Press.
- Firestone, W. A., Monfils, L., & Camilli, G. (2001). Pressure, support, and instructional change in the context of a high-stakes testing program, Paper Presented at the *American Educational Research Association*. Seattle, Washington.
- Flores, M. A., & Day, C. (2006). Contexts which shape and reshape new teachers' identities: A multi-perspective study. *Teaching and Teacher Education*, 22, 219-232.
- Foucault, M. (1972). *The archaeology of knowledge*. New York: Pantheon Books.
- Foucault, M. (1995). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans. 2nd ed.). New York: Vintage.
- Foucault, M. (1997). The ethics of concern of the self as a practice of freedom. In P. Rabinow (Ed.), *Ethics: Subjectivity and truth* (pp. 281-301). New York: The New Press.
- Fowler, F. J. (2002). *Survey research methods*. Thousand Oaks, CA: Sage.
- Franklin, C. A., & Snow-Gerono, J. L. (2007). Perceptions of teaching in an environment of standardized testing: Voices from the field. *The Researcher*, 21(1), 2-21.
- Franzak, J. K. (2004). Constructing struggling readers: Policy and the experiences of eighth-grade readers. In J. Worthy, B. Maloch, J. Hoffman, D. Schallert & C. Fairbanks (Eds.), *53rd yearbook of the National Reading Conference* (pp. 189-205). Oak Creek, WI: National Reading Conference.
- Franzosi, R. (2004). *From words to numbers: Narrative, data and social science*. Cambridge, UK: Cambridge University Press.
- Freire, P. (1995). *Pedagogy of the oppressed*. New York: Continuum.
- Freire, P., & Macedo, D. (1987). *Literacy: Reading the word and reading the world*. Westport, CT: Bergin & Garvey.
- Fullan, M. (2006). *Turnaround Leadership*. San Francisco: Jossey-Bass Publishers.

- Fullan, M. (2007). Change the terms for teacher learning. *National Staff Development*, 28(3), 35-36.
- Fullan, M., with Stiegelbauer, S. (1991). *The new meaning of educational change*. New York: Teachers College Press.
- Gardiner, M. E. (2004). Wild publics and grotesque symposiums: Habermas and Bakhtin on dialogue, everyday life and the public sphere. In N. Crossley & J. M. Roberts (Eds.), *After Habermas: New perspectives on the public sphere*. Oxford, UK: Blackwell Publishing.
- Gardner, H., Csikszentmihalyi, M., & Damon, W. (2001). *Good work: When excellence and ethics meet*. New York, NY: Basic Books.
- Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. New York: Routledge.
- Gee, J. P. (2001). Identity as an analytical lens for research in education. *Review of Research in Education*, 25, 99-125.
- Goertz, M., & Duffy, M. (2003). Mapping the landscape of high-stakes testing and accountability programs. *Theory into Practice*, 42(1), 4-11.
- Goldberg, M. F. (2004). The high-stakes test mess. *Education Digest: Essential Readings Condensed for Quick Review*, 69, 8-15.
- Goodman, K. S. (1996). *A common-sense look at the nature of language and the science of reading*. Portsmouth, NH: Heinemann.
- Goodson, I. F. (1993). The devil's bargain: Educational research and the teacher, *Education Policy Analysis* (Vol. 1, pp. 1-16).
- Gramsci, A. (1971). *Selections from prison notebooks* (T. Q. Hoare & G. N. Smith, Trans.). London: Lawrence & Wishart.
- Grant, S. G. (2000). Teachers and tests: Exploring teachers' perceptions of changes in the New York state testing program. *Educational Policy Analysis Archives*, 8(14).
- Green, J., & Winters, J. (2004). *Pushed out or pulled up? Exit exams and dropout rates in public high schools*. New York: Manhattan Institute.

- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, 2(1), 7-22.
- Groves, R. M., Dillman, D. A., Eltinge, J. L., & Little, R. J. A. (Eds.). (2002). *Survey nonresponse*. New York: John Wiley & Sons, Inc.
- Gunzenhauser, M. G. (2003). High-stakes testing and the default philosophy. *Theory into Practice*, 42(1), 51-58.
- Guthrie, J. T. (2002). Preparing students for high-stakes test taking in reading. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 370-389). Newark, Delaware: International Reading Association.
- Guthrie, J. W., & Springer, M. G. (2004). *A nation at risk* revisited: Did "wrong" reasoning result in 'right' results? At what cost? *Peabody Journal of Education*, 79(1), 7-35.
- Habermas, J. (1995). The entwinement of myth and enlightenment: Max Horheimer and Theodor Adorno (F. G. Lawrence, Trans.). In *The philosophical discourse of modernity: Twelve lectures*. Cambridge, MA: MIT Press.
- Hakim, C. (1987). *Research design: Strategies and choices in the design of social research*. London: Allen & Unwin.
- Haladyna, T., Nolen, S., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Halfpenny, P. (2001). Positivism in the 20th century. In B. Smart & G. Ritxer (Eds.), *Handbook of Social Theory* (pp. 371-385). Thousand Oaks, CA: Sage.
- Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language*. London: Edward Arnold.
- Hamilton, D., & McWilliam, E. (2001). Ex-centric voices that frame research on teaching. In V. Richardson (Ed.), *Handbook on Research on Teaching* (4th ed., pp. 17-43). Washington, D.C.: AERA.

- Hammerness, K., Darling-Hammond, L., Bransford, J., Berliner, D. C., Cochran-Smith, M., & McDonald, M. (2005). How teachers develop. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 358-389). Mahwah, NJ: Erlbaum.
- Hammersley, M. (Ed.). (1992). *Deconstructing the quantitative - qualitative divide*. Aldershot: Avebury.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41).
- Hanson, A. (2006). *No child left behind: High-stakes testing and teacher burnout in urban elementary schools*. University of Phoenix, Phoenix.
- Harre, R., & van Lagenhove, L. (Eds.). (1999). *Positioning theory: Moral contexts of intentional action*. Oxford, UK: Blackwell Publishers Ltd.
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.
- Hess, F. M. (2002). Reform, resistance ... retreat? The predictable politics of accountability in Virginia. In D. Ravitch (Ed.), *Brooking paper on education policy* (pp. 69-122). Washington: DC: Brookings Institute.
- Heubert, J. P. (2002). *Disability, race and high stakes testing*. Retrieved October 15, 2007, from http://www.cast.org/publications/ncac/ncac_disability.html
- Heubert, J. P., & Hauser, R. (Eds.). (1999). *High-stakes: Testing for tracking, promoting, and graduation*. Washington, D.C.: National Academy Press.
- Hilliard, A. G., III. . (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education*, 51(4), 293-304.
- Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.
- Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482-492.

- Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.
- Hong, W.-P., & Youngs, P. (2008). Does high-stakes testing increase cultural capital among low-income and racial minority students. *Education Policy Analysis Archives*, 16(6) [On-line], <http://epaa.asu.edu/epaa/v16n6/v16n6.pdf>.
- Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure? *Theory into Practice*, 42(1), 30-41.
- Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher*, 19, 4.
- Huber, J., & Whelan, K. (1999). A marginal story as a place of possibility: negotiating self on the professional knowledge landscape. *Teaching and Teacher Education*, 15, 381-396.
- Huberman, M. (1993). *The lives of teachers*. New York: Teachers' College Press.
- Huck, S. W. (2008). *Reading statistics and research*. Boston: Pearson.
- IRA. (1999). *High-stakes assessment in reading: A position statement of the International Reading Association*. Newark, DW: International Reading Association.
- Jaeger, R. M. (1997). Survey research methods in education. In R. M. Jaeger (Ed.), *Complementary methods for research in education* (2nd ed., pp. 449-476). Washington, DC: American Educational Research Association.
- Johnson, D. D., & Johnson, B. (2006). *High stakes: Poverty, testing, and failure in American schools* (2nd ed.). Lanham, Maryland: Rowman & Littlefield
- Johnson, S. M. (2002, July 4, 2008). What teachers need in a high-stakes world. *Harvard Graduate School of Education News*.
- Johnston, P. (2007, Saturday, December 1, 2007). *The discursive development of agency in a first grade classroom*. Paper presented at the National Reading Conference, Austin, Texas.

- Johnston, P. H. (1998). The consequences of the use of standardized tests. In S. Murphy (Ed.), *Fragile evidence: A critique of reading assessments*. Mahwah, NJ: Lawrence Erlbaum.
- Jones, B. D., & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of the high-stakes testing. *Educational Evaluation & Policy Analysis*, 12(39).
- Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), 199-203.
- Jones, M. G., Jones, B. D., & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing*. Lanham, Maryland: Rowman & Littlefield.
- Kauffman, D. (2005). Curriculum prescription and curriculum constraints: Second-year teachers' perceptions. *NGT Working Paper* [On-line], <http://www.gse.harvard.edu/~ngt>.
- Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). *The effects of standardized testing*. Boston: Kluwer-Nijhoff.
- Kellaghan, T., Madaus, G. F., & Raczek, A. E. (1996). *The use of external examinations to improve student motivation*. Washington, D.C.: American Educational Research Association.
- Kennedy, M. (2005). *Inside teaching: How classroom life undermines reform*. Cambridge, MA: Harvard University Press.
- Kincheloe, J., & Tobin, J. (2006). Doing educational research in a complex world: Preface. In K. Tobin & J. Kincheloe (Eds.), *Doing educational research: A handbook* (pp. 3-14). Rotterdam, The Netherlands: Sense Publishers.
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Kohn, A. (2002). Standardized testing: Separating the wheat children from the chaff children. In S. Ohanian (Ed.), *What happened to recess and why are our children struggling in kindergarten*. New York, NY: McGraw-Hill.

- Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996). *The perceived effects of the Maryland school performance assessment program*. Los Angeles: Center for the Study of Evaluation, University of California.
- Kozol, J. (2005). *The shame of the nation: The restoration of apartheid schooling in America*. New York: Crown Publishers.
- Krosnick, J. A., & Schuman, H. (1988). Attitude, intensity, importance, and certainty and susceptibility to response effects. *Journal of Personality and Social Psychology*, 54, 940-952.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Larson, M., & Philips, D. (2005). Becoming a teacher of literacy: The struggle between authoritative discourses. *Teaching Education*, 16(4), 311-323.
- Lasky, S. (2003). Teacher professional vulnerability in a context of standards based reforms, *Paper presented at the American Educational Research Association*. Chicago.
- Lasky, S. (2004). An exploration of teacher professional vulnerability in a context of large-scale reform. Toronto, Ontario: University of Toronto.
- Lasky, S. (2005). A sociocultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education*, 21, 899-916.
- Lather, P. (1991). *Getting smart*. New York: Routledge.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Lazear, E. P. (2006). Speeding, terrorism, and teaching to the test. *Quarterly Journal of Economics*, 121(3), 1029-1061.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(4), 4-16.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3-13.

- Lipman-Blumen, J. (1985). The creative tensions between liberal arts and specialization. *Liberal Education*, 71(1), 18.
- Litwin, M. S. (1995). *How to measure survey reliability and validity*. Thousand Oaks, California: Sage Publications.
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.
- Lowenberg Ball, D., & Forzani, F. M. (2007). What makes education research "educational"? *Educational Researcher*, 36(9), 529-540.
- Luke, C., & Gore, J. (1992). *Feminisms and critical pedagogy*. New York: Routledge.
- Madaus, G. F. (1988a). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education*, 65(3), 29-46.
- Madaus, G. F. (1988b). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum* (pp. 83-121). Chicago: University of Chicago Press.
- Madaus, G. F., West, M., Harmon, M., Lomax, R., & Viator, K. (1992). *The influence of testing on teaching math and science in grades 4-12*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Mathison, S., & Freeman, M. (2003). Constraining teachers' work: Dilemmas and paradoxes created by state mandated testing. *Education Policy Analysis Archives*, 11(34).
- Mathison, S., & Freeman, M. (2006). Teacher stress and high-stakes testing, How using one measure of academic success leads to multiple teacher stressors. In R. G. Lambert & C. J. McCarthy (Eds.), *Understanding teacher stress in an age of accountability*. Greenwich, CT: Information Age Publishing.
- McDermott, R. P. (1987). The explanation of minority failure, again. *Anthropology & Education Quarterly*, 18(4), 361-364.
- McGill-Franzen, A. (2000). Policy and Instruction: What is the relationship? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 889-908). Mahwah: NJ: Erlbaum Associates.

- McGill-Franzen, A., & Allington, R. L. (2006). Contamination of current accountability systems. *Phi Delta Kappan*, 87(10), 762-766.
- McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading Research Quarterly*, 30, 934-956.
- McNeil, L. M. (2000). *Contradictions of school reform: Education costs of standardized testing*. New York: Routledge.
- McNeil, L. S., Coppola, E., Radigan, J., & Heilig, J. V. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives*, 16(3).
- Mehan, H. (1990). The school's work of sorting students. In D. Boden & D. H. Zimmerman (Eds.), *Talk and social structure*. Cambridge: Polity Press.
- Mehrens, W. A., & Popham, W. J. (1992). How to evaluate the legal defensibility of high-stakes tests. *Applied Measurement in Education*, 5, 265-283.
- Morrow, J. (2001). *Choosing excellence: "Good enough" schools are not good enough*: Scarecrow Press.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- Mitchell, M. L., & Jolley, J. M. (2007). *Research design explained*. Belmont, CA: Thomson Wadsworth.
- Moje, E. B. (2004). Powerful spaces: Tracing the out-of-school literacy spaces of Latino/a youth. In K. Leander & M. Sheehy (Eds.), *Space matters: Assertions of space in literacy practice and research* (pp. 15-38). New York: Peter Lang.
- Muijs, D. (2004). *Doing quantitative research in education with SPSS*. Thousand Oaks, CA: Sage Publishers.
- Nardi, P. (2006a). *Doing survey research: A Guide to quantitative research* (2nd ed.). Boston: Pearson Education, Inc.

- Nardi, P. (2006b). *Interpreting data: A guide to understanding research*. Boston: Pearson.
- National Commission on Excellence. (1983). *A nation at risk: The imperative for educational reform*. Washington, D. C.: Government Printing Office.
- Neuman, W. L. (2006). *Social research methods: Qualitative and quantitative approaches*. Boston: Pearson.
- Newman, I., & Benz, C. R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale, Illinois: Southern Illinois University Press.
- Nichols, S. L., & Berliner, D. C. (2005). The inevitable corruption of indicators and educators through high-stakes testing. *Education Policy Studies Laboratory* [Online], <http://edpolicylab.org>.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, Massachusetts: Harvard Education Press.
- Noddings, N. (2007). *When school reform goes wrong*. New York: Teachers College Press.
- Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1989). *A survey of Arizona teachers and schools administrators on the uses and effects of standardized achievement testing*. Phoenix: Arizona State University, West Campus.
- Olson, L. (2002, April 24). Survey shows state testing alters instructional practices. *Education Week*, pp. 14-15.
- Pahl, K., & Rowsell, J. (2005). *Literacy and education: Understanding the new literacy studies in the classroom*. Thousand Oaks, CA: Sage.
- Paige, R. (2001). The back page: No Child Left Behind. *Carnegie Reporter*, 1(2) [Online], <http://www.carnegie.org/reporter/02/backpage/index.html>.
- Paris, S. (2000). Trojan horse in the schoolyard. *Issues in Education*, 6(1/2), 1-16.
- Paris, S., & McEvoy, A. (2002). Harmful and enduring effects of high-stakes testing. *Issues in Education*, 6(1/2), 1-16.

- Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher*, 20(5), 12-20+40.
- Patton, M. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey on teachers*. Chestnut Hill, MA: Boston College: Center for the Study of Testing, Evaluation, and Educational Policy.
- Perkinson, H. J. (1995). *The imperfect panacea: American faith in education*. Boston: McGraw-Hill
- Public Agenda. (2001). Reality Check 2001. *Education Week*, 20, S1-S8.
- Public Agenda. (2002). Reality Check 2002. *Education Week*, 21, S1-S8.
- Punch, K. F. (1998). *Introduction to social research: Quantitative and qualitative approaches*. Thousand Oaks, California: Sage Publishers.
- Ravitch, D. (1995). *National standards in American education: A citizen's guide*. Washington, DC: The Brookings Institution.
- Raymond, M. E., & Hanushek, E. A. (2003). High-stakes research. *Education Next*, 3, 48-55.
- Rea, L. M., & Parker, R. A. (2005). *Designing and conducting survey research: A comprehensive guide* (3rd ed.). San Francisco: Jossey-Bass.
- Resnick, D., & Resnick, L. (1985). Standards, curriculum, and performance: A historical and comparative perspective. *Educational Researcher*, 14(4), 5-20.
- Rex, L. A., & Nelson, M. C. (2004). How teachers' professional identities position high-stakes test preparation in their classrooms. *Teachers College Record*, 106(6), 1288-1331.

- Richardson, V. (Ed.). (1997). *Constructivist teacher education: Building a world of new understandings*. Washington, DC: The Falmer Press.
- Riddle Buly, M., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, 24(3), 219-239.
- Roderick, M., Jacob, B., & Bryk, A. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation and Policy Analysis*, 24(4), 333-357.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.
- Rosenshine, B. (2003). High-stakes testing: Another analysis. *Education Policy Archives*, 11(24).
- Rothman, R. (1995). *Measuring up: Standards, assessment, and school reform*. San Francisco, California: Jossey-Bass.
- Salant, P., & Dillman, D. A. (1994). *How to conduct your own survey*. New York: John Wiley & Sons, Inc.
- Salinas, C. (2006). Teaching in a high-stakes testing setting: What becomes of teacher knowledge. In S. G. Grant (Ed.), *Measuring history: Cases of state-level testing across the United States*. Charlotte, NC: Information Age Publishing.
- Salvaterra, M., & Adams, D. (1995). Departing from tradition: Two schools' stories. *Educational Leadership*, 53(3), 32-35.
- Sapsford, R. (2007). *Survey research* (2nd ed.). Thousand Oaks, CA: Sage.
- Schon, D. (1983). *The reflective practitioner*. New York: Basic Books.
- Schultz, B. D., & Oyler, C. (2006). We make this road as we walk together: Sharing teacher authority in a social action curriculum project. *Curriculum Inquiry*, 36(4), 423-451.
- Scott, C. (1961). Research on mail surveys. *Journal of the Royal Statistical Society*, 124, 143-205.

- Scruggs, T. E., Mastropieri, M. A., & McDuffie, K. A. (2007). Co-teaching in inclusive classrooms: A metasynthesis of qualitative research. *Exceptional Children*, 73(4), 392-416.
- Shanahan, T. (2000). Research synthesis: Making sense of the accumulation of knowledge in reading. In M. L. Kamil, P. B. Mosenthal, J. Peters & R. Barr (Eds.), *Handbook of Reading Research* (Vol. III). Mahwah, NJ: Erlbaum Associates.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shotter, J. (1989). Social accountability and the social construction of 'you'. In J. Shotter & K. Gergen (Eds.), *Texts of identity*. London: Sage.
- Sireci, S. G. (2007). On validity theory and test validation. *American Educational Research Journal*, 36(8), 477-481.
- Sirotnik, K. A. (Ed.). (2004). *Holding accountability accountable: What ought to matter in public education*. New York: Teachers College Press.
- Sleeter, C. E., & Stillman, J. (2007). Navigating accountability pressures. In C. E. Sleeter (Ed.), *Facing accountability in education: Democracy and equity at risk* (pp. 13-29). New York: Teachers College Press.
- Sloan, K. (2006). Teacher identity and agency in school worlds: Beyond the all-good/all-bad discourse on accountability-explicit curriculum policies. *Curriculum Inquiry*, 36(2), 119-152.
- Sloane, F. C., & Kelly, A. E. (2003). Issues in high-stakes testing programs. *Theory into Practice*, 42(1), 12-17.
- Smith, F. (1994). *Understanding reading: A psycholinguistic analysis of reading and learning to read* (5th ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.

- Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education*, 51(5), 334-344.
- Smith, M. L., Nobel, A. J., Heinecke, W., Seck, M., Parish, C., Cabay, M., et al. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona student assessment program (ASAP): Equity and teacher capacity building*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Smith, M. S., & O'Day, J. (1990). *Systemic school reform*. London Taylor & Francis.
- Solorzano, R. W. (2008). High-Stakes testing: Issues, Implications, and Remedies for English Language Learners. *Review of Educational Research*, 78(2), 260-329.
- Spillane, J. (2004). *Standards deviation: How schools misunderstand education policy*. Cambridge, Massachusetts: Harvard Education press.
- Spillane, J., & Thompson, C. (1997). Reconstructing notions of local capacity: The local education agency's capacity for ambitious reform. *Education Evaluation and Policy Analysis*, 19, 185-203.
- Spring, J. (2004). *American Education*. New York, NY: McGraw-Hill.
- Stecher, B., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment*, 7(4), 259-281.
- Stecher, B. M., Hamilton, L. S., & Gonzalez, G. (2003). *Working smarter to leave no child behind: Practical insights for school leaders*. Santa Monica, CA: RAND Corporation.
- Stone, D. (2002). *Policy paradox: The art of political decision making*. New York: W. W. Norton & Company.
- Strauss, A. L. (1987). *Qualitative analysis for the social sciences*. Cambridge, MA: Cambridge University Press.
- Street, B. (1984). *Literacy in theory and practice*. Cambridge: Cambridge University.
- Swanson, C. B. (2006). Making the connection: A decade of standards-based reform and achievement. *Editorial Projects in Education research*.

- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Taylor, G., Shepard, L. A., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Thomas, R. M. (2005). *High-stakes testing: Coping with collateral damage*. Mahwah, New Jersey: Lawrence Erlbaum
- Thorne, S. E., Jensen, L., Kearney, M. H., Nobli, G. W., & Sandelowski, M. (2004). Qualitative metasynthesis: reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, 14(10), 1342-1365.
- Turner, R. (2001). Role theory. In J. Turner (Ed.), *Handbook of sociological theory* (pp. 233-254). New York: Kluwer Academic/Plenum.
- Urdu, T. C., & Paris, S. G. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy*, 8(2), 137-156.
- Valenzuela, A. (Ed.). (2004). *Leaving children behind: How "Texas-style" accountability fails Latino youth*. New York: State University of New York Press.
- Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44(3), 519-558.
- Valli, L., Croninger, R. G., Chambliss, M. J., Graeber, A. O., & Buese, D. (2008). *Test driven: High-stakes accountability in elementary schools*. New York: Teachers College Press.
- Vogler, K., Novick, B., Kress, J., & Elias, M. J. (2002). The impact of high-stakes, state-mandated student performance assessment on teacher's instructional practices. *Education*, 123(1), 39-55.
- Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

- Wagner, J. (1993). Ignorance in educational research: Or, how can you "not" know that? *Educational Researcher*, 22(5), 15-23.
- Wasley, P. A. (1991). *Teachers who lead: the rhetoric of reform and the realities of practice*. New York: Teachers College Press.
- Weedon, C. (1987). *Feminist practice and poststructural theory*. New York: Blackwell.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.
- Wertsch, J. (1991). *Voices of mind: A sociocultural approach to mediated action*. Cambridge, Mass.: Harvard University Press.
- Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.
- Wolcott, H. (1994). *Transforming qualitative data*. Thousand Oaks, CA: Sage.
- Wortman, P. M. (1994). Judging research quality. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 97-110). New York: Russell Sage Foundation.
- Wright, W. E. (2002). The effects of high-stakes testing in an inner-city elementary school: The curriculum, the teachers, and the English language learners. *Current Issues in Education*, 5(5) [On-line], <http://cie.ed.asu.edu/volume5/number5>.
- Wright, W. E., & Choi, D. (2005). Voices from the classroom: A statewide survey of experienced third-grade English language learner teachers on the impact of language and high-stakes testing policies in Arizona. *Education Policy Studies Laboratory*.
- Yanow, D. (2000). *Conducting interpretive policy analysis*. Thousand Oaks, CA: Sage.

APPENDICES

**APPENDIX A: TEACHER SURVEY ON THE IMPACT OF STATE-
MANDATED TESTING PROGRAMS**

NBETPP



The National Board on Educational Testing and Public Policy

TEACHER SURVEY ON THE IMPACT OF STATE-MANDATED TESTING PROGRAMS

National Board on Educational Testing and Public Policy
Boston College

Instructions for completing the questionnaire: The purpose of this questionnaire is to gather information about state-mandated testing and its impact on classroom instruction and student learning from teachers in elementary, middle, and high schools. These issues are particularly important for the education reform efforts that are currently taking place.

A key term should be defined as it is used in the questionnaire: **state-mandated tests** are those standardized tests that your state requires of its schools at specific grade levels (such as a statewide basic skills test or a test required for high school graduation).

We are interested in your candid beliefs and practices about important issues related to these tests. Your individual responses will be kept strictly confidential and will not be provided to any other person or group. Since you have been selected as part of a national sample, your responses are extremely important if we are to represent accurately what teachers across the United States think about these issues.

If you currently teach more than one class, respond about the class with which you last met before completing this questionnaire. Think only of this class and/or subject as you complete this survey.

Some of the questions may not be relevant to what takes place in your class, school district, or state. In that case, please skip the question and continue completing the survey. Please return the completed questionnaire in the enclosed self-addressed stamped envelope within the next week. We thank you in advance for participating in this important study.

IMPORTANT MARKING INSTRUCTIONS

- Use a No. 2 pencil only.
- Do not use ink, ballpoint, or felt tip pens.
- Make solid marks that fill the circle completely.
- Erase cleanly any marks you wish to change.
- Make no stray marks on this form.
- Do not fold, tear, or mutilate this form.



WRONG MARKS
⊗ ⊗ ⊗ ⊗

RIGHT MARK

PLEASE DO NOT WRITE IN THIS AREA



- | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <input type="radio"/> AL | <input type="radio"/> HI | <input type="radio"/> MA | <input type="radio"/> NM | <input type="radio"/> SD |
| <input type="radio"/> AK | <input type="radio"/> ID | <input type="radio"/> MI | <input type="radio"/> NY | <input type="radio"/> TN |
| <input type="radio"/> AZ | <input type="radio"/> IL | <input type="radio"/> MN | <input type="radio"/> NC | <input type="radio"/> TX |
| <input type="radio"/> AR | <input type="radio"/> IN | <input type="radio"/> MS | <input type="radio"/> ND | <input type="radio"/> UT |
| <input type="radio"/> CA | <input type="radio"/> IA | <input type="radio"/> MO | <input type="radio"/> OH | <input type="radio"/> VT |
| <input type="radio"/> CO | <input type="radio"/> KS | <input type="radio"/> MT | <input type="radio"/> OK | <input type="radio"/> WA |
| <input type="radio"/> CT | <input type="radio"/> KY | <input type="radio"/> NE | <input type="radio"/> OR | <input type="radio"/> WY |
| <input type="radio"/> DE | <input type="radio"/> LA | <input type="radio"/> NV | <input type="radio"/> PA | <input type="radio"/> WV |
| <input type="radio"/> DC | <input type="radio"/> ME | <input type="radio"/> NH | <input type="radio"/> RI | <input type="radio"/> WI |
| <input type="radio"/> FL | <input type="radio"/> MD | <input type="radio"/> NJ | <input type="radio"/> SC | <input type="radio"/> WY |
| <input type="radio"/> GA | | | | |

- ☐ All (Elementary Education)
☐ English
☐ Math
☐ Science
- ☐ Social Studies
☐ Special Education
☐ Other: _____

- ☐ Kindergarten ☐ 4 ☐ 7 ☐ 10
☐ 1 ☐ 5 ☐ 8 ☐ 11
☐ 2 ☐ 6 ☐ 9 ☐ 12
☐ 3

- ☐ Urban
☐ Suburban
☐ Rural

- ☐ Above average
- ☐ Average
- ☐ Below average

1

- 1

	Strongly Disagree	Disagree	Agree	Strongly Agree
...ing because the	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...achievement as a teacher's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...g program.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...at ALL teachers in my state	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...worth the investment of time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...that any commercially	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... (CAT) measures.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...to use are compatible with	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...of education students have	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ents on the state-mandated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...thing writing because it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ents are smaller on the	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...d achievement tests (e.g.,	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ents to learn.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...cores on the	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...minority students know and	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...ts the quality of education in	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3/8" spine
per

If you teach more than one class, please respond about the class with which you last met when answering questions 52 through 59.

52. Are students placed in the class based on their achievement (i.e., tracked)? ☐ Yes ☐ No
53. Which one of the following categories best describes the ability/achievement level of this class?
- ☐ High ability or achievement ☐ Low ability or achievement
- ☐ Average ability or achievement ☐ Mixed ability or achievement
54. How many students are in this class? ☐ 1-15 ☐ 16-20 ☐ 21-25 ☐ 26-30 ☐ 31+

PLEASE DO NOT WRITE IN THIS AREA

☒ ○○○○○○○○○○○○○○○○○○○

If you teach more than one class, please respond about the class with which you last met when answering questions 55 through 59.

For questions 55 and 56, mark the response that most closely estimates the percent of your total class that each group falls into. Since a student may fall into more than one category, the percents need NOT total 100%.

	0%	1%-10%	11%-20%	21%-30%	31%-40%	41%-50%	51%-60%	61%-70%	71%-80%	81%-90%	91%-100%
55. African-American	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
American Indian or Alaskan Native	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Asian or Pacific Islander	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
White	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hispanic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
56. Limited English Proficiency (LEP)/English as a Second Language (ESL)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Special Education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

57. Which response best describes the socio-economic status (SES) of most of the students in each group? Please mark the appropriate column for each group of students.

Groups

Students in your class
Students in your school
Students in your district

	Mostly low SES	Mostly middle SES	Mostly high SES	Mixture of all three
Students in your class	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students in your school	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students in your district	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

58. What percent of your students ...

have a computer at home?
prefer to write first drafts using a computer?
can keyboard moderately well (20 words per minute or more)?

	0%	1%-10%	11%-20%	21%-30%	31%-40%	41%-50%	51%-60%	61%-70%	71%-80%	81%-90%	91%-100%
have a computer at home?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
prefer to write first drafts using a computer?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
can keyboard moderately well (20 words per minute or more)?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

59. About how often do you typically use assessments of the following types?

Multiple-choice questions
Open response (short answer)
Extended response (essay)
Performance assessment (e.g., debates, experiments, portfolios)
Group work yielding an individual product
Group work yielding a group product

	Never	Once or twice a year	Three to six times a year	Once a month	Two to three times a month	Once a week or more
Multiple-choice questions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open response (short answer)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Extended response (essay)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Performance assessment (e.g., debates, experiments, portfolios)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Group work yielding an individual product	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Group work yielding a group product	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

60. How do you prepare your students for your state-mandated test? Mark ALL that apply:

- ☐ I do no special test preparation.
- ☐ I teach test-taking skills.
- ☐ I encourage students to work hard and prepare.
- ☐ I provide rewards for test completion.
- ☐ I teach the standards or frameworks known to be on the test.
- ☐ I provide students with items similar to those on the test.
- ☐ I provide test-specific preparation materials developed commercially or by the state.
- ☐ I provide students with released items from the state-mandated test.

3/8" scale
per inch

61. The following is a list of ways in which state-mandated test results are used. For each item please indicate how appropriate you feel the specific use is. Please rate each use with the following scale.

Place students in gifted and talented/honors programs
 Place students in special education
 Remediate students
 Promote or retain students in grade
 Graduate students from high school
 Group students by ability in a grade
 Evaluate teacher or administrator performance
 Award teachers or administrators financial bonuses
 Reward schools financially
 Hold schools accountable
 Hold the district accountable
 Rank schools publicly
 Fire faculty/staff
 Award school accreditation
 Place public schools in receivership/state takeover
 Evaluate charter schools
 Evaluate voucher programs

Very Inappropriate	Moderately Inappropriate	Moderately Appropriate	Very Appropriate
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

62. In what ways, if any, has the amount of time spent on each of the following activities changed in your school in order to prepare students for the state-mandated testing program? Please skip to question #63 if you are a first year teacher.

Instruction in tested areas
 Instruction in areas not covered by the state-mandated test
 Instruction in tested areas with high stakes attached (e.g., promotion, graduation, teacher rewards)
 Instruction in tested areas without high stakes attached
 Instruction in the fine arts
 Instruction in physical education
 Instruction in foreign language
 Instruction in industrial/vocational education
 Student free time (e.g., recess, lunch)
 Field trips (e.g., museum tour, hospital tour)
 Class trips (e.g., circus, amusement park)
 Student choice time (e.g., games, computer work)
 Organized play (e.g., games with other classes)
 Enrichment school assemblies (e.g., professional choral group performances)
 Administrative school assemblies (e.g., awards ceremonies)
 Classroom enrichment activities (e.g., guest speakers)
 Student performance (e.g., class plays)
 Parental contact

Increased a Great Deal	Moderately Increased	Stayed About the Same	Moderately Decreased	Decreased a Great Deal
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

236371-2/2

- | Yes | No |
|-----------------------|-----------------------|
| <input type="radio"/> | <input type="radio"/> |
| <input type="radio"/> | <input type="radio"/> |
| <input type="radio"/> | <input type="radio"/> |
| <input type="radio"/> | <input type="radio"/> |
| <input type="radio"/> | <input type="radio"/> |

70. How often do your own **STUDENT'S** results on the state-mandated test influence your teaching?
Mark only one response.
- ☐ Daily
 - ☐ A few times a week
 - ☐ A few times a month
 - ☐ A few times a year
 - ☐ Never
 - ☐ I did not receive students' test results in time to use them
 - ☐ I teach a grade and/or subject that does not receive students' test results
 - ☐ I teach a grade and/or subject that should get students' results but did not receive them

71. Using the following scale, indicate the extent of your agreement with the statements below regarding the state-mandated test results.

The *individual* student reports on test performance are easy to interpret.
 The *individual* student reports on test performance provide useful information.
 The *school* reports on student performance are easy to interpret.
 The *school* reports on student performance provide useful information.
 The *district* reports on student performance are easy to interpret.
 The *district* reports on student performance provide useful information.

Have Never Seen the Report	Strongly Disagree	Disagree	Agree	Strongly Agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

72. Do YOU use the results of the state-mandated test for any of the following activities? Mark ALL that apply.

- | | |
|--|---|
| <input type="radio"/> Group students within my class | <input type="radio"/> Give feedback to students |
| <input type="radio"/> Evaluate student progress | <input type="radio"/> Give feedback to parents |
| <input type="radio"/> Assess my teaching effectiveness | <input type="radio"/> Determine student grades (in whole or in part) |
| <input type="radio"/> Select instructional materials | <input type="radio"/> Do not get the results back in time to use them |
| <input type="radio"/> Plan my instruction | <input type="radio"/> None of the above |
| <input type="radio"/> Plan curriculum | |

73. Are the results from the state-mandated test used in your DISTRICT to make decisions about the following?

Please mark ALL that apply.

- | | |
|---|---|
| <input type="radio"/> Place students in gifted and talented/honors programs | <input type="radio"/> Hold the district accountable |
| <input type="radio"/> Place students in special education | <input type="radio"/> Rank schools publicly |
| <input type="radio"/> Remediate students | <input type="radio"/> Fire faculty/staff |
| <input type="radio"/> Promote or retain students in grade | <input type="radio"/> Award school accreditation |
| <input type="radio"/> Graduate students from high school | <input type="radio"/> Place public schools in receivership/state takeover |
| <input type="radio"/> Group students by ability in a grade | <input type="radio"/> Evaluate charter schools |
| <input type="radio"/> Evaluate teacher or administrator performance | <input type="radio"/> Evaluate voucher programs |
| <input type="radio"/> Award teachers or administrators financial bonuses | <input type="radio"/> None of the above |
| <input type="radio"/> Reward schools financially | <input type="radio"/> Other: _____ |
| <input type="radio"/> Hold schools accountable | |

74. Is there at least one person at your school that teachers can turn to for accurate information about the state-mandated testing program?

☐ Yes ☐ No

75. How adequate has professional development in the following areas been in preparing teachers in your district to implement the state-mandated testing program?

Knowledge of state curriculum standards or frameworks
 Alignment of the classroom curriculum to the state curriculum standards/frameworks
 Alignment of the classroom curriculum to the state-mandated test
 Test preparation strategies
 Administration of the state-mandated test
 Interpretation of the test results
 Use of test results

No Professional Development	Very Inadequate	Inadequate	Adequate	Very Adequate
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The figure shows the distribution of responses for four statements. Each statement has a corresponding horizontal bar divided into five segments representing different levels of agreement. The segments are labeled from left to right as Strongly Disagree, Disagree, Neutral, Agree, and Strongly Agree. The approximate percentages for each segment are as follows:

Statement	Strongly Disagree (%)	Disagree (%)	Neutral (%)	Agree (%)	Strongly Agree (%)
1. I am confident about my ability to manage my own business.	10	20	30	30	10
2. I have enough resources to start my business.	10	20	30	30	10
3. I have enough knowledge to start my business.	10	20	30	30	10
4. I have enough experience to start my business.	10	20	30	30	10

- ## Background Information

☐ African American
☐ American Indian or Alaskan Native
☐ Asian
☐ White
☐ Pacific Islander
☐ Hispanic
☐ Other, please specify: _____

[illegible]

Mark Reflex® by NCS EM-236371-1:654321 ED99 Printed in U.S.A. © Copyright 2000 by National Computer Systems, Inc. All rights reserved.

☒ ○

**APPENDIX B: TEACHER SURVEY ON THE EFFECTS OF HIGH-
STAKES TESTING ON TEACHER PEDAGOGY, PRACTICE AND
IDENTITY**

Teacher Survey on the Effects of High-Stakes Testing on Teacher Pedagogy, Practice and Identity

The purpose of this survey is to gather information about the effects of high-stakes testing on teacher pedagogy, practice and identity. High-stakes tests are those tests which are mandated at specific grade levels and hold specific consequences and impact your work as a teacher. I am interested in your candid beliefs and practices about important issues related to these tests. You have been chosen to participate as part of an Eastern Tennessee sample of grades 3 – 8 teachers. Completion of the survey will require 20 - 25 minutes of your time. Identities of the teacher participants will be protected. The researcher will not attempt to pursue any individual teacher identifying information in reporting and summarizing data. No reference will be made in oral or written reports that could link participants to the study. If you have any questions or would like to comment on other aspects of the study, please refer to the Invitation Letter. Thank you for your time and for helping me with this research.

COUNTY

In what school county do you currently teach?

- ☐ Alcoa County
- ☐ Anderson County
- ☐ Blount County
- ☐ Knox County
- ☐ Roane County
- ☐ Sevier County

GRADES

What grade level(s) do you currently teach?

- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8

- ☐ 9
☐ 10
☐ other : _____

SUBJECTS

What subject(s) do you teach?

SPECIALTY_AREA

Do you teach in any specialty area?

- ☐ Yes
☐ No

SPECIALTY_AREA_DESC

Please list the area of specialty.

LOCATION

Which category best describes your school?

- ☐ Urban
☐ Suburban
☐ Rural

COMPARE

How do your school's results on the state-mandated test compare to those of other schools in your state?

- ☐ Above average
☐ Average
☐ Below average

TITLE

Is your school a Title 1 school?

- ☐ Yes
☐ No

AYP

Is your school meeting Adequate Yearly Progress (AYP)?

- ☐ Yes
☐ No

SCALE

Please indicate the extent to which you agree with each of the following statements.

	Stongly Disagree	Disagree	Agree	Stongly Agree
Teachers in my school do NOT use computers when teaching writing because the state-mandated writing test is handwritten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The state-mandated test is compatible with my daily instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The state-mandated test is as accurate a measure of student achievement as a teacher's judgement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My district's curriculum is aligned with the state-mandated testing program.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The state-mandated test is based on a curriculum framework that ALL teachers in my state should follow.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall, the benefits of the state-mandated testing program are worth the investment of time and money.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
What the state-mandated test measures is about the same as what any commercially available standardized achievement test (e.g., Stanford 9, ITBS, CAT) measures.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teacher morale is high in	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

my school.

The instructional texts and materials that the district requires me to use are compatible with the state-mandated test.

☐ ☐ ☐ ☐

Scores on the state-mandated test accurately reflect the quality of education students have received.

☐ ☐ ☐ ☐

The state-mandated testing program is just another fad.

☐ ☐ ☐ ☐

Teachers have high expectations for the performance of all students on the state-mandated test.

☐ ☐ ☐ ☐

My school's (district's) policy forbids using computers when teaching writing because it does NOT match the format of the state-mandated writing test.

☐ ☐ ☐ ☐

Performance differences between minority and nonminority students are smaller on the state-mandated test than on commercially available standardized achievement tests (e.g., Stanford 9, ITBS, CAT).

☐ ☐ ☐ ☐

The state-mandated test motivates previously unmotivated students to learn.

☐ ☐ ☐ ☐

Teachers feel pressure from the district superintendent to raise scores on the state-mandated test.

☐ ☐ ☐ ☐

The state-mandated test is NOT an accurate measure of what minority students know and can do.

☐ ☐ ☐ ☐

Media coverage of state-mandated test results accurately reflects the quality of education in my state.

☐ ☐ ☐ ☐

Many low scoring students will do better on the state-mandated test if they receive specific preparation for it.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student morale is high in my school.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
If I teach to the state standards or frameworks, students will do well on the state-mandated test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Many students in my class feel that, no matter how hard they try, they will still do poorly on the state-mandated test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The state-mandated test measures high standards of achievement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Media coverage of state-mandated testing issues has been unfair to teachers.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The state-mandated test is NOT an accurate measure of what students who are acquiring English as a second language know and can do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The majority of my students try their best on the state-mandated test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Many students are extremely anxious about taking the state-mandated test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teachers have high expectations for the in-class academic performance of students in my school.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Differences among schools on the state-mandated test are more a reflection of students' background	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

characteristics than of
school effectiveness.

My school has an
atmosphere conducive to learning. ☐ ☐ ☐ ☐

Teachers feel pressure from
parents to raise scores on
the state-mandated test. ☐ ☐ ☐ ☐

Media coverage of state-
mandated testing issues
adequately reflects the
complexity of teaching. ☐ ☐ ☐ ☐

There is so much pressure
for high scores on the state-
mandated test that teachers
have little time to teach
anything not on the test. ☐ ☐ ☐ ☐

The state-mandated test
has brought much needed
attention to education
issues in my district. ☐ ☐ ☐ ☐

Students are under intense
pressure to perform well on
the state-mandated test. ☐ ☐ ☐ ☐

My tests are in the same
format as the state-
mandated test. ☐ ☐ ☐ ☐

Teachers in my school want
to transfer out of the grades
where the state-mandated
test is administered. ☐ ☐ ☐ ☐

The state-mandated testing
program leads some
teachers in my school to
teach in ways that
contradict their own ideas of
good educational practice. ☐ ☐ ☐ ☐

Teachers in my school have
found ways to raise state-
mandated test scores
without really improving
student learning. ☐ ☐ ☐ ☐

State-mandated testing has
caused many students in
my district to drop out of
high school. ☐ ☐ ☐ ☐

Teachers feel pressure from
the building principal to
raise scores on the state-
mandated test. ☐ ☐ ☐ ☐

State-mandated test results have led to many students being retained in grade in my district.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Administrators in my school believe students' state-mandated test scores reflect the quality of teachers' instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My tests have the same content as the state-mandated test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Many students in my school cheat on the state-mandated test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

PREP

**How do you prepare your students for your state-mandated test?
Mark all that apply.**

- ☐ I do no special test preparation
- ☐ I teach test-taking skills
- ☐ I encourage students to work hard and prepare
- ☐ I provide rewards for test completion
- ☐ I teach the standards or frameworks known to be on the test
- ☐ I provide students with items similar to those on the test
- ☐ I provide test-specific preparation materials developed commercially or by the state
- ☐ I provide students with released items from the state-mandated test

RESULT_USE

The following is a list of ways in which state-mandated test results are used. For each item please indicate how appropriate you feel the specific use is. Please use the following scale.

	Very inappropriate	Moderately inappropriate	Moderately appropriate	Very appropriate
Place students in gifted and talented/honors programs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Place students in special education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Remediate students	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Promote or retain students in grade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Graduate students from high school	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluate teacher or	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

administrator performance

Award teachers or administrators financial bonuses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reward school financially	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rank school publicly	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fire faculty/staff	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Award school accreditation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Place public schools in receivership/state takeover	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluate charter school	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluate voucher programs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIRST_YEAR

Are you a first year teacher?

- ☐ Yes
☐ No

CHANGE

In what ways, if any, has the amount of time spent on each of the following activities changed in your school in order to prepare students for the state-mandated testing program?

Please skip to the next question if you are a first year teacher.

	Increased a great deal	Moderately increased	Stayed about the same	Moderately decreased	Decreased a great deal
Instruction in tested areas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instruction in areas not covered by the state-mandated test	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instruction in tested areas with high stakes attached (e.g., promotion, graduation, teacher rewards)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instruction in tested areas without high stakes attached	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instruction in the fine arts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instruction in physical education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instruction in foreign					

language	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instruction in industrial/vocational education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student free time (e.g., recess, lunch)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Field trips (e.g., museum tour, hospital tour)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Class trips (e.g., circus, amusement park)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student choice time (e.g., professional choral group performances)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Administrative school assemblies (e.g., award ceremonies)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Classroom enrichment activities (e.g., guest speakers)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student performance (e.g., class plays)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Parental contact	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

HOURS

Approximately how many class hours PER YEAR do you spend preparing students specifically for the state-mandated test (e.g., teaching test-taking skills)?

- ☐ None
☐ 1-10
☐ 11-20
☐ 21-30
☐ More than 30

PREPARATION

When were most of the test preparation activities you conducted specifically for the state-mandated test carried out?

- ☐ No specific preparation
☐ The day before
☐ Throughout the week before
☐ Throughout the two weeks before
☐ Throughout the month before
☐ Throughout the year

TESTPREPARATION

How similar is the content of the test preparation materials you use to the content of the state-mandated test?

- ☐ Very similar
- ☐ Somewhat similar
- ☐ Somewhat dissimilar
- ☐ Very dissimilar

TEST_PREP

One test preparation strategy is to target specific groups of students. Please mark ALL that apply related to your state-mandated test.

- ☐ I do not target test preparation at specific groups of students
- ☐ I target test preparation at LEP or ESL students
- ☐ I target test preparation at Special Education (SPED) students
- ☐ I target test preparation at students on the border of passing the state-mandated test.
- ☐ I target test preparation at students who are on the border of moving to the next performance level.

TESTPREPARATION2

Have you heard of any of the following activities taking place during the state-mandated test administration at your school?

Have any of the teachers in your school done any of the following?

	Yes	No
Given students hints about answers	<input type="radio"/>	<input type="radio"/>
Pointed out mismarked items to students	<input type="radio"/>	<input type="radio"/>
Given some students more than the allowed time	<input type="radio"/>	<input type="radio"/>
Provided instruction during the test	<input type="radio"/>	<input type="radio"/>
Changed student answers on the test	<input type="radio"/>	<input type="radio"/>

RELY

Does your school rely on any of the following strategies to influence students to do their best work on the state-mandated test? Please mark ALL that apply.

- ☐ Discussing the importance to the school of good performance on the test
- ☐ Holding student assemblies to motivate students
- ☐ Publicly recognizing students for good performance
- ☐ Scheduling special activities (e.g., pizza parties, field trips)
- ☐ Providing free time as a reward to students

- ☐ Linking performance to eligibility for participation in extra curricular activities (e.g., athletics, clubs)
- ☐ Giving prizes to reward students
- ☐ Requiring/recommending summer school
- ☐ Retaining students in grade
- ☐ Using scores for assigning report card grades
- ☐ Placing students in classes (e.g., honors, remedial)
- ☐ Exempting students who do well from required course work

ALIGNMENT

How often do your SCHOOL'S results on the state-mandated test influence your own teaching? Mark only one response.

- ☐ Daily
- ☐ A few times a week
- ☐ A few times a month
- ☐ A few times a year
- ☐ Never
- ☐ I did not receive the school's test results in time to use them
- ☐ I teach a grade and/or subject that does not receive the school's test results
- ☐ I teach a grade and/or subject that should get results but did not receive them

ALIGN

Using the following scale, indicate the extent of your agreement with the statements below regarding the state-mandated test results

	Strongly Disagree	Disagree	Agree	Strongly Agree	Have Never Seen the Report
The individual student reports on test performance are easy to interpret	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The individual student reports on test performance provide useful information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The school reports on student performance are easy to interpret	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The school reports on student performance provide useful information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The district reports on					

student performance are easy to interpret ☐ ☐ ☐ ☐ ☐

The district reports on student performance provide useful information ☐ ☐ ☐ ☐ ☐

INFLUENCE2

Your state-mandated testing program influence the amount of time you spend on ...

	Stongly Disagree	Disagree	Agree	Strongly Agree
Whole group instruction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Critical thinking skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Individual seat work	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Basic skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Students working together in small groups (cooperative learning)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Concept development using manipulatives or experiments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Problems that are likely to appear on the state-mandated test	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

YEARS_EXP

How many years of teaching do you have, including this year?

(0 - 255)

--	--	--

EDUCATION

What is the highest level of education that you hold?

- ☐ Bachelor
- ☐ Masters
- ☐ PhD
- ☐ Other : _____

CERTIFICATION

Please list any additional certification or licensures

AGE3

What is your age?

(0 - 255)

--	--	--

RACE

Select the category that best describes you.

- ☐ African American
- ☐ American Indian or Alaskan Native
- ☐ Asian
- ☐ White
- ☐ Pacific Islander
- ☐ Hispanic
- ☐ Other : _____

IDENTITY

If you would like to offer any additional comments about the effects of high-stakes testing in relation to your work and identity as a teacher please write them here in the space provided.

EMAIL

Thank you for taking the time to complete the survey. If you would like to enter your email address to be included in a random drawing please click on the tab below. Each respondent completing the survey and adding their name to the draw will be eligible to be sent 1 of 55 teacher professional books. *Email addresses will never be used in conjunction with the data; all contact information will be electronically separated from the data.*

FOUCSGROUP

The researcher is interested in setting a further research agenda to follow the completion of this study. Would you be interested in participating in a teacher focus group session or a single teacher interview session planned for late November 2008?

- ☐ Yes
- ☐ No
- ☐ No Answer

RESEARCHINQUIRY

Please enter an email address or phone number where the researcher may contact you about participating.

APPENDIX C: INITIAL FACTOR LOADINGS

Initial Factor Loadings

	Component			
	1	2	3	4
Students are under intense pressure to perform well on the state-mandated test.	.638			
There is so much pressure for high scores on the state-mandated test that teachers have little time to teach anything not on the test.	.583			
Teachers feel pressure from the building principal to raise scores on the state-mandated test.	.569		.247	-.314
Many students are extremely anxious about taking the state-mandated test.	.550			
Teachers in my school want to transfer out of the grades where the state-mandated test is administered.	.549			
State-mandated testing has caused many students in my district to drop out of high school.	.514			
The state-mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good educational practice.	.507	-.297		
Teachers feel pressure from the district superintendent to raise scores on the state-mandated test.	.491			
The state-mandated test is NOT an accurate measure of what students who are acquiring English as a second language know and can do.	.484	-.365		
Teachers feel pressure from parents to raise scores on the state-mandated test.	.446	.214		
The state-mandated test is NOT an accurate measure of what minority students know and can do.	.427	-.370		
Media coverage of state-mandated testing issues has been unfair to teachers.	.422	-.307		
Administrators in my school believe students' state-mandated test scores reflect the quality of teachers' instruction.	.397			-.277
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	.388			
'Many students in my class feel that, no matter how hard they try, they will still do poorly on the state-mandated test.'	.346			-.202
The state-mandated testing program is just another fad.	.244		-.233	
Teachers in my school do NOT use computers when teaching writing because the state-mandated writing test is handwritten.	.241			-.223
Scores on the state-mandated test accurately reflect the quality of education students have received.		.589	.356	
The state-mandated test motivates previously unmotivated students to learn.	-.259	.540		
Media coverage of state-mandated test results accurately reflects the quality of education in my state.		.514		
'Overall, the benefits of the state-mandated testing program are worth the investment of time and money.'	-.265	.499	.394	
The state-mandated test is as accurate a measure of student achievement as a teacher's judgement.	-.211	.492	.380	

The state-mandated test measures high standards of achievement.	.476	.208
The state-mandated test has brought much needed attention to education issues in my district.	.470	
State-mandated test results have led to many students being retained in grade in my district.	.218	.467
Media coverage of state-mandated testing issues adequately reflects the complexity of teaching.	.405	
Many low scoring students will do better on the state-mandated test if they receive specific preparation for it.	.377	
Differences among schools on the state-mandated test are more a reflection of students' background characteristics than of school effectiveness.	.321	-.372
My school's (district's) policy forbids using computers when teaching writing because it does NOT match the format of the state-mandated writing test.	.372	-.243
'Performance differences between minority and nonminority students are smaller on the state-mandated test than on commercially available standardized achievement tests (e.g., Stanford 9, ITBS, CAT).'	.369	
'If I teach to the state standards or frameworks, students will do well on the state-mandated test.'	.349	.279
Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness.	-.343	
Many students in my school cheat on the state-mandated test.	.323	-.299
The state-mandated test is compatible with my daily instruction.		.694
The state-mandated test is based on a curriculum framework that ALL teachers in my state should follow.		.600
My tests have the same content as the state-mandated test.		.598
My district's curriculum is aligned with the state-mandated testing program.		.555
The instructional texts and materials that the district requires me to use are compatible with the state-mandated test.		.486
My tests are in the same format as the state-mandated test.		.394
'What the state-mandated test measures is about the same as what any commercially available standardized achievement test (e.g., Stanford 9, ITBS, CAT) measures.'		.323
My school has an atmosphere conducive to learning.		.756
Teachers have high expectations for the in-class academic performance of students in my school.	.216	.641
Student morale is high in my school.	.209	.621
Teacher morale is high in my school.	-.229	.231
Teachers have high expectations for the performance of all students on the state-mandated test.		.224
The majority of my students try their best on the state-mandated test.		.425

Loadings below .200 not included.

APPENDIX D: FINAL FACTOR LOADINGS

Final Factor Loadings

	Component			
	1	2	3	4
Scores on the state-mandated test accurately reflect the quality of education students have received.	.654			
'Overall, the benefits of the state-mandated testing program are worth the investment of time and money.'	.581			
The state-mandated test is as accurate a measure of student achievement as a teacher's judgement.	.577			
Media coverage of state-mandated test results accurately reflects the quality of education in my state.	.553			
The state-mandated test motivates previously unmotivated students to learn.	.542			
The state-mandated test has brought much needed attention to education issues in my district.	.499			
Differences among schools on the state-mandated test are more a reflection of students' background characteristics than of school effectiveness.	-.475			
The state-mandated test measures high standards of achievement.	.467			
Media coverage of state-mandated testing issues adequately reflects the complexity of teaching.	.450			
State-mandated test results have led to many students being retained in grade in my district.	.428			
Score differences from year to year on the state-mandated test reflect changes in the characteristics of students rather than changes in school effectiveness.	-.396			
Media coverage of state-mandated testing issues has been unfair to teachers.	-.374			
'If I teach to the state standards or frameworks, students will do well on the state-mandated test.'	.362			
Many low scoring students will do better on the state-mandated test if they receive specific preparation for it.	.358			
Students are under intense pressure to perform well on the state-mandated test.		.670		
Teachers feel pressure from the building principal to raise scores on the state-mandated test.		.570		
Many students are extremely anxious about taking the state-mandated test.		.561		
There is so much pressure for high scores on the state-mandated test that teachers have little time to teach anything not on the test.		.553		

Teachers in my school want to transfer out of the grades where the state-mandated test is administered.	.523
State-mandated testing has caused many students in my district to drop out of high school.	.514
Teachers feel pressure from the district superintendent to raise scores on the state-mandated test.	.494
Teachers feel pressure from parents to raise scores on the state-mandated test.	.492
The state-mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good educational practice.	.485
Administrators in my school believe students' state-mandated test scores reflect the quality of teachers' instruction.	.435
Teachers in my school have found ways to raise state-mandated test scores without really improving student learning.	.371
'Many students in my class feel that, no matter how hard they try, they will still do poorly on the state-mandated test.'	.336
My school has an atmosphere conducive to learning.	.780
Teachers have high expectations for the in-class academic performance of students in my school.	.654
Student morale is high in my school.	.634
Teacher morale is high in my school.	.611
Teachers have high expectations for the performance of all students on the state-mandated test.	.556
The majority of my students try their best on the state-mandated test.	.431
The state-mandated test is compatible with my daily instruction.	.689
My tests have the same content as the state-mandated test.	.642
My district's curriculum is aligned with the state-mandated testing program.	.574
The state-mandated test is based on a curriculum framework that ALL teachers in my state should follow.	.568
The instructional texts and materials that the district requires me to use are compatible with the state-mandated test.	.489
My tests are in the same format as the state-mandated test.	.412

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

VITA

Jan E. Blake earned her PhD. in Literacy Studies with a concentration in Reading at the University of Tennessee, December 2008. She holds a Bachelor of Education degree from the University of Victoria in Secondary Education: Geography and Visual Arts, and a Masters of Education degree from the University of British Columbia in Language and Literacy: Reading. Dr. Blake has taught grades K -12 in the public schools, as well as undergraduate and graduate courses at the university level. As a practicing classroom teacher, special educator, early reading intervention teacher, gifted program coordinator, primary support teacher, collaborative learning coordinator, Dr. Blake has brought a broad and rich education background to her doctoral studies. Dr. Blake has written three Canadian Provincial Curriculums, provided extensive and varied workshops and programmed professional development for teachers, administrators and parents at the local, district and provincial levels. Dr. Blake is active with several professional organizations. While at UTK, Dr. Blake has worked with the State Improvement Grant for Tennessee, supervised Education 100 Apple Corps students, provided professional development workshops to Tennessee teachers, and acted as a Field Representative of the National Reading Conference for Tennessee. Dr. Blake's research interests include: striving readers, assessment and evaluation, and teacher identity.